



## Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom.

Pierre Fillard, Maxime Descoteaux, Alvina Goh, Sylvain Gouttard, Ben Jeurissen, James Malcolm, Alonso Ramirez-Manzanares, Marco Reisert, Ken Sakaie, Fatima Tensaouti, et al.

### ► To cite this version:

Pierre Fillard, Maxime Descoteaux, Alvina Goh, Sylvain Gouttard, Ben Jeurissen, et al.. Quantitative evaluation of 10 tractography algorithms on a realistic diffusion MR phantom.. *NeuroImage*, 2011, 56 (1), pp.220-34. 10.1016/j.neuroimage.2011.01.032 . inria-00559191

**HAL Id: inria-00559191**

**<https://inria.hal.science/inria-00559191>**

Submitted on 25 Jan 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Quantitative Evaluation of 10 Tractography Algorithms on a Realistic Diffusion MR Phantom

Pierre Fillard<sup>1,2</sup>, Maxime Descoteaux<sup>3</sup>, Alvina Goh<sup>4</sup>,  
Sylvain Gouttard<sup>5</sup>, Ben Jeurissen<sup>6</sup>, James Malcolm<sup>7</sup>,  
Alonso Ramirez-Manzanares<sup>8</sup>, Marco Reisert<sup>9</sup>, Ken Sakaie<sup>10</sup>,  
Fatima Tensaouti<sup>11</sup>, Ting Yo<sup>12</sup>, Jean-François Mangin<sup>2</sup>, Cyril Poupon<sup>2</sup>

January 25, 2011

<sup>1</sup>Parietal Research Team, INRIA Saclay Île-de-France, Neurospin, France

<sup>2</sup>Laboratory of Computer-Assisted Neuro-Imaging, CEA Saclay, Neurospin, France

<sup>3</sup>MOIVRE Center, Computer Science department, Université de Sherbrooke, Canada

<sup>4</sup>Department of Mathematics, National University of Singapore, Singapore

<sup>5</sup>Scientific Computing and Imaging Institute, University of Utah, USA

<sup>6</sup>IBBT-VisionLab, Department of Physics, University of Antwerp, Belgium

<sup>7</sup>Psychiatry Neuroimaging Laboratory, Brigham and Womens Hospital, Harvard Medical School, USA

<sup>8</sup>Mathematics Department, University of Guanajuato, Mexico

<sup>9</sup>Department of Radiology, Medical Physics, University Hospital Freiburg, Germany

<sup>10</sup>Imaging Institute, The Cleveland Clinic, Cleveland, USA

<sup>11</sup>National Institute for Medical Research, INSERM, U825, France

<sup>12</sup>Max Planck Institute for Human Cognitive and Brain Sciences, Germany

## Abstract

As it provides the only method for mapping white matter fibers *in vivo*, diffusion MRI tractography is gaining importance in clinical and neuroscience research. However, despite the increasing availability of different diffusion models and tractography algorithms, it remains unclear under different imaging parameters, how would one make the optimal choice for a fiber reconstruction method. Consequently, it is of utmost importance to have a quantitative comparison of these models and algorithms and a deeper understanding of the corresponding strengths and weaknesses. In this work, we use a common dataset with known ground truth and a reproducible methodology to quantitatively evaluate the performance of various diffusion models and tractography algorithms. To examine a wide range of methods, the dataset, but not ground truth, was released to the public for evaluation in a contest, the "Fiber Cup". 10 fiber reconstruction methods were evaluated. The results provide evidence that: 1. For high SNR datasets, diffusion models such as (fiber) orientation distribution functions correctly model the underlying fiber distribution and can be used in conjunction with streamline tractography, and 2. For medium or low SNR datasets, a prior on the spatial smoothness of either the diffusion model or the fibers is recommended for correct modelling of the fiber distribution and proper tractography results. The phantom dataset, the ground truth fibers, the evaluation methodology and the results obtained so far will remain publicly available on: <http://www.lnao.fr/spip.php?rubrique79> to serve as a comparison basis for existing or new tractography methods. New results can be submitted to [fibercup09@gmail.com](mailto:fibercup09@gmail.com) and updates will be published on the webpage.

# 1 Introduction

The unique ability of Diffusion Weighted MRI (DW-MRI) based fiber tractography to map, *in vivo*, the architecture of white matter pathways has ignited strong interest in clinical and neuroscience research. Applications include improved assessment of a range of neurological and psychiatric disorders [Ciccarelli et al., 2008] and characterization of anatomical connections [Johansen-Berg and Rushworth, 2009]. The potential of tractography to help map anatomical connections played a significant role in motivating an ambitious project to map the human "connectome"<sup>1</sup>.

Despite the increasing availability of different diffusion models and tractography algorithms, it remains unclear under different imaging parameters how one would make the optimal choice for a fiber reconstruction method. Notably, one objective of these methods is to model and track in the presence of complex fiber configurations such as crossings or kissings, but objective comparison of the performance of each has not been done yet. Consequently, it is of utmost importance to have a quantitative comparison of these models and algorithms and a deeper understanding of the corresponding strengths and weaknesses. To emphasize the need for such comparative study, we briefly review the diffusion models and tractography algorithms available in the literature.

The role of the diffusion model is to characterize the underlying fiber distribution at each voxel, indicating the number and orientation of distinct fiber compartments, and many choices are available. The most commonly used diffusion model is the diffusion tensor (DT) [Basser et al., 1994] as estimated from Diffusion Tensor Imaging (DTI). DTI is popular due to the simplicity of the model and of the imaging acquisition, which requires as few as six diffusion weighted images (DWI) and is compatible with clinical conditions. However, the DT can only characterize one fiber compartment per voxel – a problem in regions of crossing fibers where at least two are expected. Consequently, several alternatives have been proposed to overcome this limitation. Most of them are based on High Angular Resolution Diffusion Imaging (HARDI) [Tuch, 2002], which uses several tens to a few hundreds of DWI. An incomplete list includes the Q-ball model [Tuch, 2004], the composite hindered and restricted model of diffusion (CHARMED) [Assaf and Basser, 2005], the diffusion orientation transform (DOT) [Özarslan et al., 2006], multi-tensor distributions [Tuch, 2002, Liu et al., 2004], the orientation distribution function (ODF) [Tuch, 2004], the fiber ODF using spherical deconvolution (SD) [Tournier et al., 2004, Tournier et al., 2007, Descoteaux et al., 2009], the ball & stick model [Behrens et al., 2003], the mixtures of Wisharts [Jian and Vemuri, 2007], and the persistent angular structure MRI (PAS-MRI) [Jansons and Alexander, 2003].

A large number of tractography algorithms have been developed to map fibers through the entire brain based on information from the voxel-level diffusion models. The algorithms can be categorized into two main classes: deterministic and probabilistic. Deterministic algorithms follow the main fiber directions as revealed by the diffusion model and generate sequences of points that are considered as fibers. Probabilistic algorithms repeat many times the deterministic version by randomly

---

<sup>1</sup><http://www.humanconnectomeproject.org>

perturbing the main fiber directions each time, and produce maps of connectivity. Such maps indicate the probability that a given voxel is connected to a reference position. A small sampling of tractography algorithms indicates the dizzying array of choices. Among deterministic tractography algorithms, streamline algorithms were developed first [Mori et al., 1999b, Conturo et al., 1999, Basser et al., 2000], followed by more elaborated tensor deflection algorithms [Weinstein et al., 1999, Lazar et al., 2003] or more global approaches [Poupon et al., 2001, Mangin et al., 2002]. This list is not exhaustive and many other DT-based algorithms exist. HARDI-based techniques include the generalization of streamline tracking to use ODFs computed from diffusion spectrum imaging [Tuch, 2002, Hagmann et al., 2004], multi-tensor tracking [Kreher et al., 2005, Bergmann et al., 2007], the q-ball tracking [Chao et al., 2007b], and ODF tracking [Campbell et al., 2006, Descoteaux et al., 2009]. Probabilistic tractography methods include DT-based algorithms [Parker et al., 2003, Behrens et al., 2003, Lazar and Alexander, 2005, Friman et al., 2006, Ramirez-Manzanares and Rivera, 2006, Savadjiev et al., 2008, Koch et al., 2002, Zhang et al., 2009], calculation of geodesics in a DT-warped space [Lenglet, 2006, Jbabdi et al., 2004], and numerous HARDI-based methods [Parker and Alexander, 2005, Perrin et al., 2005a, Seunarine et al., 2006, Behrens et al., 2007b, Jbabdi et al., 2007, Savadjiev et al., 2008, Chao et al., 2007a, Seunarine et al., 2007, Haroon and Parker, 2007, Kaden et al., 2007, Jeurissen et al., 2010]. Once again, this list is not exhaustive.

The increasing number of diffusion models and tractography algorithms is both a blessing and a curse: While each diffusion model has the ability of modeling more complex diffusivity and each tractography algorithm has the capability of obtaining previously undetected fiber tracts, it is becoming obscure to decide which diffusion model in conjunction to which tractography algorithm should be used for a particular application. More important, the performance of each method may vary and one cannot ensure that the reconstructed fibers are effectively representative of the true fiber organization, or if they are artifacts produced by method.

Lack of publicly available ground truth for validation makes it difficult to objectively choose among the huge variety of diffusion models, tractography algorithms, and combinations thereof. A number of validation studies have been performed, but each has its drawbacks. Campbell et al [Campbell et al., 2006] used two rat spinal cords to create a biological phantom. Unfortunately, the fiber configuration was sparse as compared with the brain. Other synthetic phantoms were proposed, for instance by Perrin et al [Perrin et al., 2005b] (two bundles crossing at 90 degrees), and Pullens et al. [Pullens et al., 2010], or by Moussavi et al. [Moussavi-Biugui et al., 2010] (two thick spherical bundles crossing at either 45 or 90 degrees). However, all of them lack the complex crossing and kissing fibers found in the brain. Recently, Fieremans et al. [Fieremans et al., 2008] proposed the design of a realistic diffusion MR phantom made of complex structures similar to the white matter geometry such as fiber crossings and curved fibers, which is challenging for tractography algorithms. Heat-shrinking tubes were used to pack and compress fibers to increase anisotropy. Complex 3D geometries can therefore be designed. Furthermore, Pullens et al. [Pullens et al., 2010] showed that fiber density and orientation could be controlled using such design strategy. Nevertheless, the proposed phantoms exhibit only simple crossings, although having a rather realistic 3D geometry. In addition,

attempts at validating diffusion MRI with histological data have been presented by Leergaard et al [Leergaard et al., 2010] and Anderson et al [Anderson et al., 2006]. However, they are difficult to reproduce for non-experts in histology and only address validation of diffusion models at the voxel level, not larger scale tractography. Identically, using neuronal tracers as done by Dauguet et al. [Dauguet et al., 2007] is hardly generalizable to a large number of fiber bundles. By contrast, numerical simulations as in [Hall and Alexander, 2009] have well-defined ground truths at the price of an oversimplification of the physics of diffusion in tissues. Moreover, they often discard acquisition artifacts (such as noise and distortion), which are important parameters to take into account in practice.

The objectives of this study are to provide a qualitative and quantitative comparison of several tractography methods on the same realistic dataset with known ground truth and to freely distribute this dataset along with the evaluation methodology so that new methods can be easily evaluated and compared to existing ones. We therefore constructed a realistic diffusion MR phantom containing numerous crossing, kissing, splitting and bending configurations that we purposely developed to this end [Poupon et al., 2010]. To compare as many different approaches as possible, we organized a tractography contest, the *Fiber Cup*, during the last Medical Image Computing and Computer Assisted Intervention (MICCAI) conference in London in September 2009. Common datasets with known ground truth were distributed publicly and researchers were invited to apply their tractography methodologies to these datasets. Contestants were blinded to ground truth. Results were analyzed and ranked based on several metrics. The contest format addresses several difficulties associated with comparing diffusion model and tractography methods. A huge number of methods have been published, and quite a few software implementations have been made publicly available for comparison. A list of the most popular diffusion MRI processing softwares includes FSL [Behrens et al., 2007a], Trackvis [Wang et al., 2007], Camino [Cook et al., 2006], 3D Slicer [Pieper et al., 2006], Brainvisa [Rivière et al., 2009], MedINRIA [Toussaint et al., 2007], ExploreDTI [Leemans et al., 2009] and MRTrix [Tournier et al., 2007]. While a single-site comparison similar to that performed by Klein et al. [Klein et al., 2009] for registration algorithms is feasible, we opted for the reverse situation in order to obtain as many contributions as possible, including some of the most recent algorithms whose implementation may not be available in a public package yet. Moreover, unlike Klein et al. who evaluated registration packages out-of-the-box without fine parameter tuning, we preferred to extract the best from each method for a fair evaluation. It means that the expertise of an algorithm’s author was required to run it properly.

The rest of the paper is organized as follows. In Section 2, we present the ground-truth dataset specifically designed for the contest followed by the evaluation methodology. In Section 3, we review the results by summarizing the 10 received contributions before detailing their qualitative and quantitative evaluation. Finally, in Section 4 we discuss about the pros and cons of each method and give some recommendations.

## 2 Material and Methods

In this section, we will first briefly review the construction of a realistic diffusion MR ground truth dataset. We will then detail the rules of the tractography competition and the methodology developed to quantitatively evaluate and compare contributions.

### 2.1 Construction of a Ground Truth Dataset

#### 2.1.1 Design of a Realistic Diffusion MR Phantom

The construction of diffusion phantoms is a challenging task involving the following two steps [Poupon et al., 2008]:

- Design of a realistic and practically feasible fiber configuration.
- Mechanical conception: Construction of a frame and container to hold the fibers under high tension in water. An adequate recipe to fill the container with a MRI compatible solution is also required.

In what follows, we depict those two steps that lead to the elaboration of our realistic diffusion MR phantom.

**Fiber configuration:** First, the desired fiber configuration must be designed. The configuration should be as realistic possible - containing crossing and kissing fibers as well as bundles of different curvatures. However, the configuration must be practically feasible. In particular, fibers lie in-plane because they have to be squeezed in between two solid dies to ensure high density and diffusion anisotropy approaching that of tissue [Poupon et al., 2008]. Consequently, complex 3D configurations are not permitted. Moreover, bundles are stretched while being squeezed to ensure fibers take a straight, smooth, trajectory within the phantom. Therefore, the bundles enter and leave the phantom, forbidding closed trajectories.

We opted for a configuration simulating a coronal section of the human brain, containing several fiber crossing and kissing configurations with different curvatures (see Fig. 1 left). The phantom comprises seven distinct bundles, and contains 3 crossings, 1 kissing, and 3 bundles that split. For a fast text referencing, those regions were numbered from 1 to 7 on Fig. 1 right. Several curvatures were also used. In particular, a U-shape bundle (Fig. 1 right, region 7) with high curvature was placed to challenge methods with a strong prior on the straightness of the bundles to achieve a perfect score: methods assuming very straight fibers are likely to fail in reconstructing this U-shape structure.

**Mechanical conception and manufacturing:** To create large bundles, hydrophobic acrylic fibers whose diameter is of the same order as the diameter of myelinated axons were used ( $20\ \mu m$ ). Those fibers are very common and can be found in any specialized clothing shops under the name "acrylic fibers for Boutis and Patchwork". A polyurethane negative and positive prints of the target bundles were manufactured and used to strongly tighten the fibers together. The diameter of the polyurethane device is of 14 cm. The phantom was filled to ensure that the same amount of fibers is found everywhere, including in the crossings. Fibers were carefully positioned such that they rigorously follow the pathways sketched

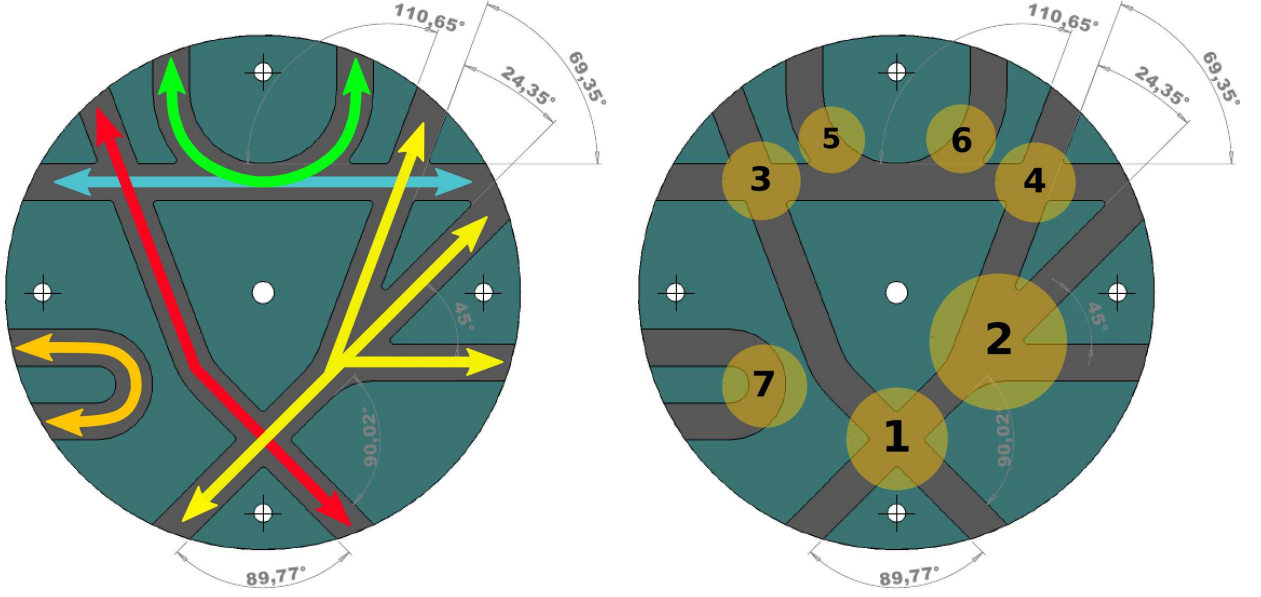


Figure 1: **Sketch of the phantom that mimics a coronal section of the human brain.** **Left:** Fiber pathways are highlighted in colors. Arrows indicate the directions of the synthetic fiber bundles. **Right:** The various crossing, splitting and kissing fiber configurations are numbered for a fast text referencing. Note that angles between crossing fibers were carefully determined, although not used during evaluation.

in Figure 1. Bundles of about 100 fibers were used to progressively fill it. First, a layer of bundles was placed everywhere in the phantom. Then, a second layer was placed everywhere except at the intersections. Without this operation, intersections would have contained twice or three times more fibers than the branches, depending on how many bundles are crossing. Finally, this process was reiterated until the desired number of fibers was positioned.

The next step consists in compressing the fibers to increase their density. Compression is carefully controlled to make sure that fibers are captured in  $1\text{mm}^2$  cross-section everywhere throughout the phantom. Practically, the positive and negative prints were squeezed while keeping fibers strongly tightened until the openings (i.e, where the fibers enter/leave the phantom) are exactly 1cm thick. As they are distributed all over the phantom, we assume that the level of compression is the same everywhere. Since we know the number of fibers and the space they are captured in, we can deduce the density of fibers, which was close to  $1900\text{ fibers/mm}^2$  everywhere, including in the crossings.

A cylindrical container compatible with MRI head coil antennas was designed to hold the phantom in the magnet. This container was made up of a plexiglass cylinder ended at its extremities by two plastic caps, equipped with taps to fill it. As standard DW-MRI uses the ultra-fast echoplanar acquisition scheme that is extremely sensitive to phase inhomogeneities, developing a filling process insuring the absence of any air bubbles in the container is of upmost importance. To do so, a dedicated platform was designed that enables a preliminary degassing of the solution, and a filling under vacuum conditions. A ultrasound beam is finally used to destroy any remaining air bubbles. The container was filled using pure distilled



water without any contrast agent.

### 2.1.2 Diffusion-Weighted MRI Acquisitions

Diffusion-weighted data of the phantom were acquired on the 3T Tim Trio MRI systems of the NeuroSpin centre, equipped with a whole body gradient coil (40 mT/m, 200 T/m/s), and using a 12-channel receive only head coil, in combination with the whole body transmit coil of the MRI system.

A single-shot diffusion-weighted twice refocused spin echo echoplanar pulse sequence was used to perform the acquisitions, while compensating for the first order Eddy currents. Two datasets were acquired at two different spatial resolutions: 3 mm isotropic and 6 mm isotropic.

Parameters for the 3 mm isotropic acquisition were as follows: field of view FOV=19.2cm, matrix 64x64, slice thickness TH=3mm, read bandwidth RBW=1775 Hz/pixel, partial Fourier factor 6/8, parallel reduction factor GRAPPA=2, repetition time TR=5s, 2 repetitions. Three diffusion sensitizations at b-values  $b=650/1500/2000$  s/mm<sup>2</sup> corresponding to the echo times TE=77/94/102 ms respectively. 3 slices were acquired. A SNR of 15.8 was measured for the baseline ( $b=0$ ) image. SNR of DWI at b-values 650/1500/2000 were evaluated at respectively 9.1/2.6/1.1<sup>2</sup>. ADC and FA values of the 3x3x3 acquisition are reported in Fig. 2 for the three b-values.

Parameters for the 6 mm isotropic acquisition were as follows: field of view FOV=38.4cm, matrix 64x64, slice thickness TH=6mm, read bandwidth RBW=1775 Hz/pixel, partial Fourier factor 6/8, parallel reduction factor GRAPPA=2, repetition time TR=5s, 1 repetition. Three diffusion sensitization's at b-values  $b=650/1500/2650$  s/mm<sup>2</sup> corresponding to the echo times TE=77/94/110 ms respectively, making a total of 6 different datasets. 1 slice was acquired. A SNR of 22.6 was measured for the baseline ( $b=0$ ) image. SNR of DWI at b-values 650/1500/2650 were evaluated at respectively 18.9/17.6/4.5. ADC and FA values of the 6x6x6 acquisition are reported in Fig. 3 for the three b-values.

The diffusion sensitization was applied along a set of 64 orientations uniformly distributed over the sphere. Note that b-values were chosen such that phantom ADC compares to brain ADC at b-values of respectively 2000, 4000 and 6000 s/mm<sup>2</sup>.

### 2.1.3 Estimation of a Ground Truth Dataset

To facilitate the evaluation of the different results submitted during the contest, we chose to restrict the analysis to a set of 16 fibers traversing 16 manually identified voxels, or seeds. The choice of those 16 spatial positions was made to ensure that a single fiber bundle passes through each of them to avoid ambiguity on the result and to facilitate the evaluation. Indeed, receiving multiple fibers per seed would have been cumbersome: the exact number of bundles crossing in a voxel is *a priori* unknown (we recall that the ground truth was not revealed to competitors). Comparison of an undetermined number of fibers to the ground truth, whose number of bundles per voxel is precisely known, is non-trivial. Consequently, we chose to

---

<sup>2</sup>We define SNR as the ratio between the signal magnitude and the noise power (i.e., standard deviation) [Kaufman et al., 1989]. We used voxels in the center of the phantom branches (see Figure 1) to estimate the signal magnitude, and voxels outside the phantom to estimate the noise power. Note that the SNR of the DWI was averaged over all gradients.

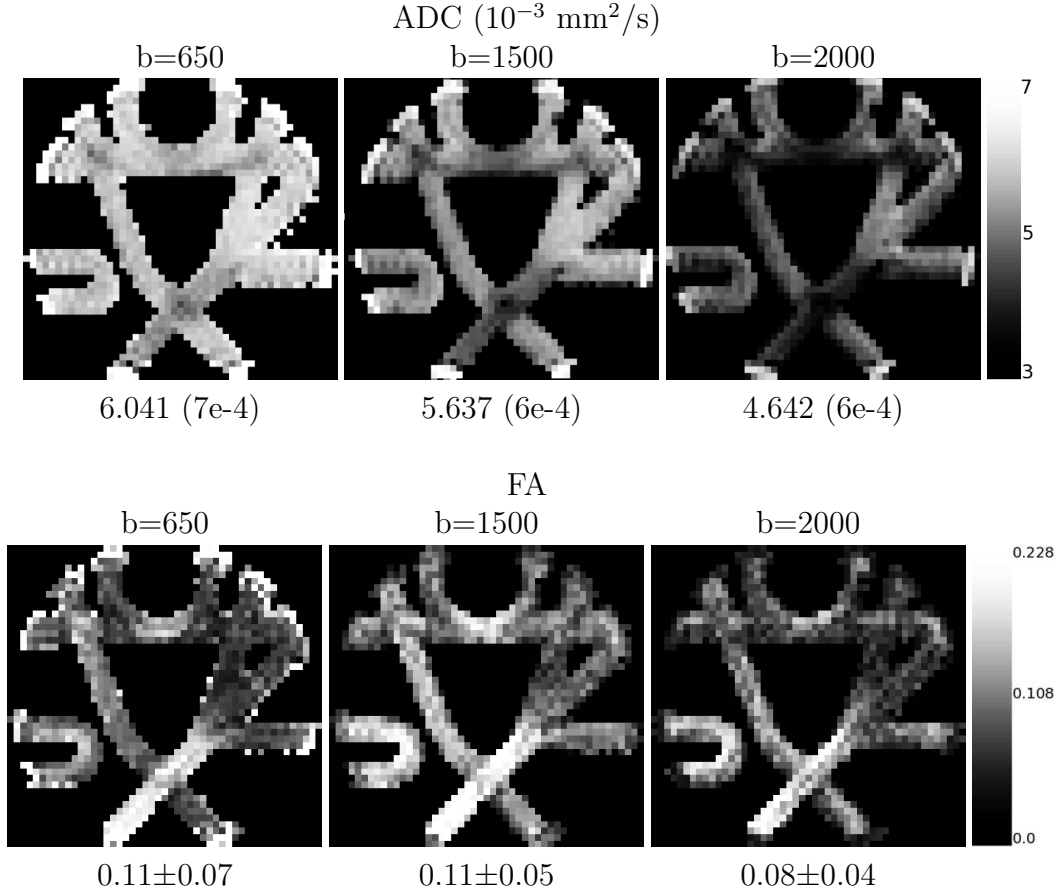


Figure 2: **ADC and FA images of the 3x3x3 phantom dataset.** **Top:** The middle slice of the ADC image is shown for the three b-values used for acquisition. The values reported are the mean (standard deviation) of the ADC values of the phantom. **Bottom:** The middle slice of the FA image is shown for the three b-values used for acquisition. The values reported are the mean (standard deviation) of the FA values of the phantom.

define seeds in voxels where a unique fiber bundle is expected. Competitors were asked to return a single representative fiber of the bundle traversing each seed voxel. The 16 chosen seeds are shown in Figure 4. They were labeled from S1 to S16. Only S9 is ambiguous since two solutions are possible, but this ambiguity was detected soon after the contest started and could not be removed.

For each seed, an estimate of the fiber bundle traversing it was manually drawn on top of the  $3 \times 3 \times 3 \text{ mm } b_0$  image by following the internal fibrous structure that was put in place during the phantom elaboration. Then, lines were smoothed using approximating cubic b-spline to remove any sampling noise, and fibers were resampled using 1000 uniformly distributed points, which formed the ground truth (Fig. 4 c). Ground truth for the  $6 \times 6 \times 6 \text{ mm}$  dataset was obtained by first registering affinely the  $3 \times 3 \times 3 \text{ mm } b_0$  image onto the  $6 \times 6 \times 6 \text{ mm } b_0$  image, and second by applying the obtained affine matrix to the previously defined fiber coordinates (Fig. 4 d).

Obviously, the nature of the ground truth itself prevents the inclusion of probabilistic tractography algorithms into the evaluation panel, since those output gener-

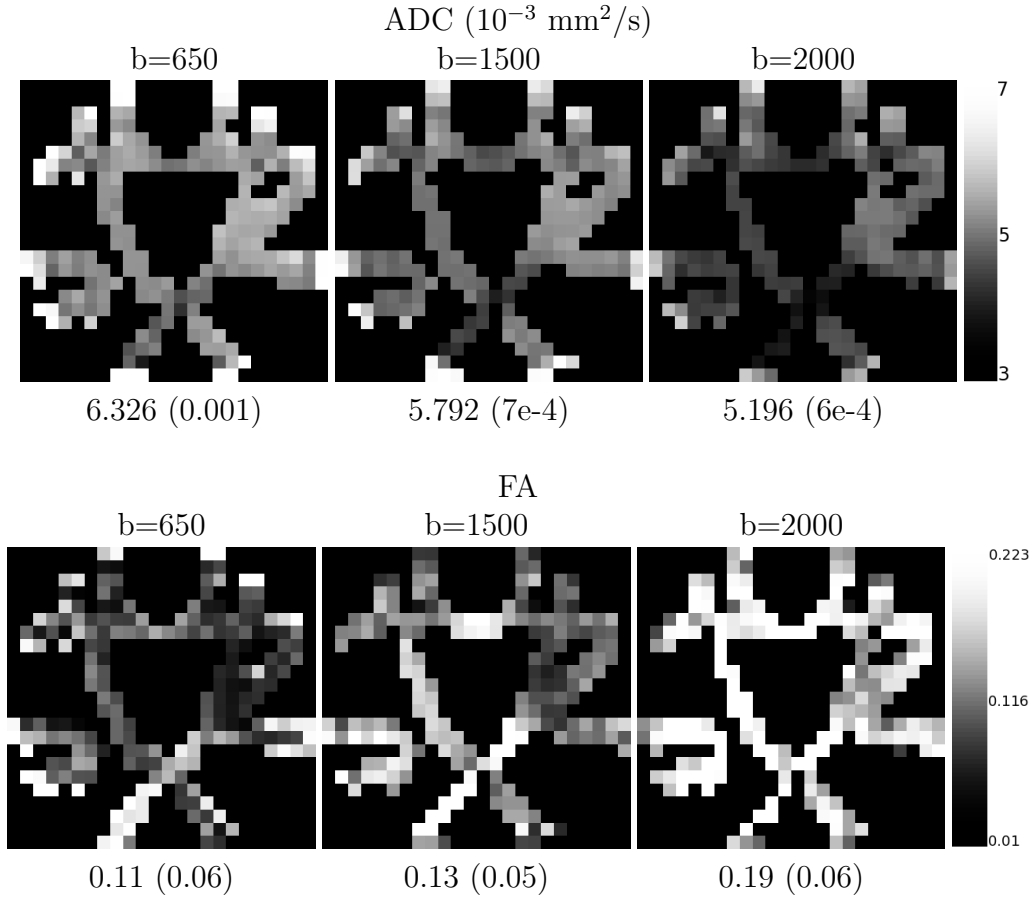


Figure 3: **ADC and FA images of the 6x6x6 phantom dataset.** **Top:** The single slice of the ADC image is shown for the three b-values used for acquisition. The values reported are the mean (standard deviation) of the ADC values of the phantom. **Bottom:** The single slice of the FA image is shown for the three b-values used for acquisition. The values reported are the mean (standard deviation) of the FA values of the phantom.

ally connectivity maps (CM) and not fiber pathways. Intuitively, defining a ground truth for such class of algorithm is still possible given the fact that we know the fiber density at every position of the phantom: if we assume that water molecules diffuse less in region of high fiber density (where diffusion space is reduced), we could define ground truth connectivity maps for every seed. However, this simplistic view does not fully reflect reality. First, there is no evidence that water molecules diffuse proportionally to fiber density: physics need to be carefully investigated (software phantoms simulating the physics of diffusion could shed light on this). Second, several other factors may impact diffusion, such as inhomogeneities of fiber density, fiber diameter or tissue permeability. For these reasons, we chose to exclude probabilistic tractography algorithms from the evaluation.

## 2.2 Contest Rules

Participants were free to use any combination of tools and algorithms that lead to the best result. DWI pre-processing was allowed. Any combination of diffusion

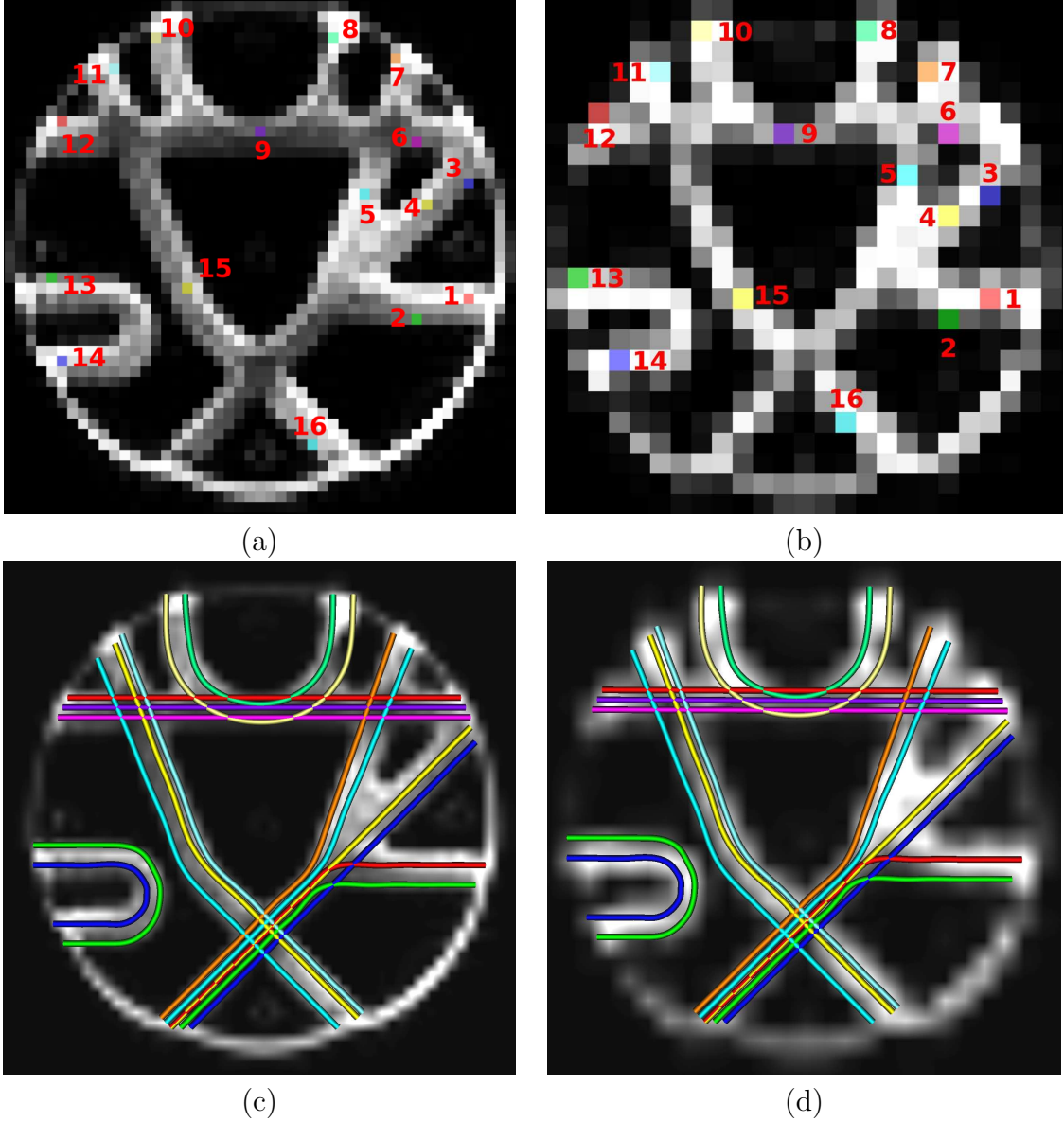


Figure 4: **The 16 seed voxels chosen for the contest.** **Top:** Seeds defined on the  $3 \times 3 \times 3$  mm (a) and  $6 \times 6 \times 6$  mm (b) datasets. **Bottom:** Ground truth fibers for both datasets. For each figure, the  $b_0$  image is shown.

models and tractography algorithms was permitted. The only restriction was that the method should be automatic and reproducible. Obviously, manual drawing of the fibers was not allowed.

Participants had only access to the raw DWI datasets. We recall that 6 datasets were available: three b-values at the 3 mm isotropic resolution, and three b-values at the 6 mm isotropic resolution. Participants were free to use any, or several, of these datasets. The 64 diffusion gradients used during acquisition were provided as a text file. The 16 seeds were also available for both 3 mm and 6 mm datasets as an image of region of interests, where the grey value of a voxel indicates its label (label 0 identified the background).

A unique representative fiber of the bundle traversing each seed was required. Each fiber has to come in a separate file named with the seed label it originates. Thus, comparison with the ground truth was easy as the file name directly points to the corresponding ground truth fiber. This design favors deterministic tracking algorithms and this issue will be discussed later in the conclusions.

An agreement on a coordinate system for expressing fiber coordinates had to be made. Moreover, the choice of a common fiber file format had to be done. Furthermore, fiber sampling can greatly differ from one participant to another and could impact the evaluation. In the following, we detail the solutions we deployed to solve those technical yet crucial issues for a fair evaluation of the contest results.

### 2.2.1 Common Fiber File Format

Due to the increasing availability of tractography softwares (DtiStudio, Brainvisa, TrackVis, MedINRIA, Slicer to quote just a few), and by extension to the existence of numerous fiber file formats, one could not reasonably rely on one of them, mainly because they can be quite complex to produce, especially for those who are not familiar with them. Instead, we chose to rely on the simplest existing format: the text file. Participants were asked to return a single text file per fiber, where the fiber coordinates are listed in sequential order (i.e., x y z coordinates of the first point, x y z coordinates of the second point, etc.), one point per line. Thus, the number of lines corresponds exactly to the number of points of a fiber.

### 2.2.2 Coordinate System

We imposed results to be expressed in physical coordinates. The transformation matrix to go from voxel to physical coordinates was included in the DWI headers (in nifti file format), and was provided to participants as such:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 3 \end{bmatrix} \times \begin{pmatrix} i \\ j \\ k \end{pmatrix} + \begin{bmatrix} 93 \\ 93 \\ -1.5 \end{bmatrix},$$

for the  $3 \times 3 \times 3$  mm and:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} -6 & 0 & 0 \\ 0 & -6 & 0 \\ 0 & 0 & 6 \end{bmatrix} \times \begin{pmatrix} i \\ j \\ k \end{pmatrix} + \begin{bmatrix} 186 \\ 186 \\ 3 \end{bmatrix},$$

for the  $6 \times 6 \times 6$  mm.  $(x, y, z)$  are real-world coordinates and  $(i, j, k)$  are voxel coordinates.

### 2.2.3 Pre-processing

Another important issue to take care of when evaluating tractography results from different participants is the fiber sampling. The sampling is very likely to differ from one submission to another: some methods produce highly sampled fibers with several hundreds or thousands of points, while others only provide a dozen of points. To normalize this, fibers were parametrized by interpolating cubic b-splines. Interpolation was chosen in order not to alter the fiber coordinates as returned by the

participants. Finally, a uniform sampling of a 1000 points matching the one used for the ground truth was performed for every fiber.

In the next section, we present the evaluation methodology used to compare tractography results with the ground truth.

## 2.3 Evaluation Methodology

Evaluation was performed on a per-fiber basis. We recall that each participant had to return a dataset composed of 16 candidate fibers matching 16 ground truth fibers. Thus, the candidate fiber passing through seed  $N$  can be compared to the ground truth fiber going through the same seed. Consequently, the evaluation methodology narrows down to the evaluation of differences between pairs of curves.

In the following, we describe the evaluation measures of curve matching we used for this contest. In particular, we show how to compute scores that evaluate not only the spatial matching of the curves, but also compare their trajectories and smoothness.

### 2.3.1 Generic Score of Fiber Match

The optimal result is realized when the candidate fiber perfectly matches the ground truth, i.e., when both fibers are superimposed. Thus, we chose to rely on the point-based Root Mean Square Error (RMSE) between the candidate fiber and the corresponding ground truth:

$$\text{RMSE}(f_1, f_2) = \sqrt{\int_0^1 \text{dist}^2(f_1(s), f_2(c(s))) ds}, \quad (1)$$

where  $f_1$  and  $f_2$  are the two fibers being compared (we recall that fibers are parametrized with b-splines so they can be expressed as functions of their arc length),  $s$  the arc length in the range  $[0, 1]$ ,  $c$  a function giving for each arc length  $s$  of  $f_1$  the corresponding arc length of  $f_2$ , and  $\text{dist}$  a metric measuring how similar the points  $f_1(s)$  and  $f_2(c(s))$  are.

The RMSE relies on a correspondence function  $c$  which is responsible for providing the arc length of fiber  $f_2$  which corresponds to arc length  $s$  on fiber  $f_1$ . The choice of  $c$  is obviously not unique, and without any prior knowledge there is no best choice for it. Without any prior on point correspondences, we chose to associate fiber points which are the closest spatially, similarly to the work [Fillard et al., 2007] on sulcal lines. Consequently, we search for the correspondence function  $c$  such that:

$$c = \min_c \int_0^1 \|f_1(s) - f_2(c(s))\|^2 ds.$$

Practically,  $c$  is obtained as follows. A distance matrix is built where each line gives the distance from a point of the first fiber to every sample points of the second. Then, following [Fillard et al., 2007], dynamic programming is used to determine the path of minimal cost within this distance matrix, which gives us the final correspondences between the arc length of both fibers. This procedure ensures that the function  $c$  is monotonically increasing, i.e., if  $s_1 \geq s_2$ ,  $c(s_1) \geq c(s_2)$ , which

ensures that two consecutive points of a fiber are associated to two other consecutive points.

Nonetheless, the mapping  $c$  is not guaranteed to be symmetric, i.e., if we denote by  $c_1$  (resp.  $c_2$ ) the mapping such that  $f_1(s)$  (resp.  $f_2(s)$ ) corresponds to  $f_2(c_1(s))$  (resp.  $f_1(c_2(s))$ ), nothing ensures that  $c_2 = c_1^{-1}$ . The consequence is that the RMSE of Eq. 1 is not symmetric:  $\text{RMSE}(f_1(s), f_2(s)) \neq \text{RMSE}(f_2(s), f_1(s))$ . As there is no real justification for this asymmetry, we chose to define a symmetrized version of the RMSE, denoted by sRMSE (for symmetric Root Mean Square Error), as the average between  $\text{RMSE}(f_1, f_2)$  and  $\text{RMSE}(f_2, f_1)$ :

$$\text{sRMSE}(f_1, f_2)(s) = \frac{1}{2} [\text{RMSE}(f_1(s), f_2(s)) + \text{RMSE}(f_2(s), f_1(s))]$$

Finally, Eq. 1 can be reformulated into a symmetric version:

$$\text{sRMSE}(f_1, f_2) = \frac{1}{2} \left( \sqrt{\int_0^1 \text{dist}^2(f_1(s), f_2(c_1(s))) ds} + \sqrt{\int_0^1 \text{dist}^2(f_2(s), f_1(c_2(s))) ds} \right) \quad (2)$$

The sRMSE depends on the chosen metric *dist*. For instance, taking the  $L_2$  norm for *dist* will result in high (resp. low) values when fibers are distant from (resp. close to) each other. On the contrary, by taking the angular difference between tangents, the sRMSE will be low (resp. high) when fibers are parallel (resp. orthogonal). Thus, by changing the metric, it is possible to evaluate how curves matches not only in terms of spatial positions but also trajectories and shape. In the following, we express three metrics that were used for the contest: the spatial metric, the tangent metric and the curve metric.

### 2.3.2 The Contest Metrics

**The Spatial Metric:** The spatial metric is simply the  $L_2$  norm between two corresponding fiber positions. If we denote by  $p_1$  and  $p_2$  two spatial positions, one can express it as:

$$\text{dist}(p_1, p_2) = \sqrt{\|p_2 - p_1\|^2} = \sqrt{(p_2 - p_1)^\top (p_2 - p_1)}. \quad (3)$$

The sRMSE endowed with the spatial metric is expressed in *mm* and ranges from 0 (overlapping points) to infinity.

**The Tangent Metric:** Let  $v_1 = f'_1(s)/\|f'_1(s)\|$  (resp.  $v_2 = f'_2(s)/\|f'_2(s)\|$ ).  $v_1$  and  $v_2$  are normalized tangent vectors to fiber points. We define the tangent metric as:

$$\text{dist}(v_1, v_2) = \left| \arccos(|v_1^\top v_2|) \frac{180}{\pi} \right| \quad (4)$$

The sRMSE endowed with the tangent metric is expressed in degrees and varies from  $0^\circ$  (parallel fibers) to  $90^\circ$  (orthogonal fibers).

**The Curve Metric:** The curvature at any position of a curve is given by:  $\kappa(f) = \frac{\|f' \times f''\|}{\|f'\|^3}$ . The curve metric is expressed as the absolute difference of the curvature between two fiber points:

$$\text{dist}(\kappa_1, \kappa_2) = |\kappa_2 - \kappa_1| \quad (5)$$

The sRMSE endowed with the curve metric is expressed in  $mm^{-1}$  and ranges from 0 to infinity.

Note that the b-spline representation of the fibers gives us an analytical expression of the first and second order derivatives necessary to evaluate the three metrics.

In the next section, we present the results of the qualitative and quantitative evaluation of the 10 contributions received during the Fiber Cup.

### 3 Results

A total of 9 individual submissions were received, including one with 2 results, making a total of 10 tractography results. Results were analyzed following the methodology described in the previous section. Computation of quantitative metrics was performed on a regular PC (Intel Core 2 Duo, 2Gb of memory).

For the sake of completeness, we also included the result of the probabilistic tractography algorithm implemented in FSL [Behrens et al., 2007a], as this is one of most widely used algorithm within the neuroscience community. While probabilistic tractography does not comply with the requirements of our quantitative evaluation methodology, its qualitative evaluation can still inform its capabilities to infer the correct fiber bundles given a seed region. Moreover, it shows that the phantom dataset can be used to evaluate probabilistic tractography algorithms. Results are presented in the Supplementary Section 1.

In this section, we first summarize the 10 contributions in terms of diffusion model and tractography algorithm chosen. Second, we give a qualitative overview of the results before presenting their in-depth quantitative analysis.

#### 3.1 Summary of Contributions

An overview of the 10 tractography methods evaluated during the contest is given in Table 1. Practical details for results reproducibility are given in the Supplementary Section 2. Even with a rather low number of contributions, we obtained a good sampling of the various diffusion model and tractography algorithms available in the literature. More precisely, the choice of the diffusion model appeared as more important than the tractography algorithm itself, often reduced to a streamline approach, although some variety can be noted.

One remark is that all participants except number 2 chose the 3 mm dataset. The choice of the higher resolution / reduced SNR dataset is interesting since a common problem in real acquisitions is to know how much of the SNR should be sacrificed in favor of the spatial resolution. A b-value of 1500 was preferably chosen, very probably to mimic the real-case scenario where a b-value of about 1000 is often used.



Table 1: **Overview of the 10 contributions.** Details about the diffusion model, tractography method and chosen dataset are given for each method. ODF-SH (resp. FOD-SH) stands for ODF (resp. FOD) parameterized with spherical harmonics. MoG stands for mixture of Gaussian. Details for method reproducibility are given in Supplementary Section 2.

Id	Diffusion Model	Tractography	Dataset	Remarks	References	Software	Contact
1	1- or 2-DT	Streamline with propagation direction following tensor with eigenvector the closest to the current direction	$3 \times 3 \times 3$ , $b = 1500$	The 2-tensor model is dropped for single tensor in voxels where the largest fiber compartment is twice as important as the smallest.	[Ramirez-Manzanares et al., 2009, Ramirez-Manzanares et al., 2007]	Matlab package	Alonso Ramirez-Manzanares
2	FOD-SH	Streamline with propagation direction following the FOD peak closest to previous direction	$6 \times 6 \times 6$ , $b = 2650$	An anisotropic denoising pre-processing step is used before FOD estimation. A non-negativity constraint ensures the FOD to be positive.	[Jeurissen et al., 2009, Jeurissen et al., 2010]	ExploreDTI	Ben Jeurissen
3	Single-DT	Streamline with propagation direction following tensor principal direction of diffusion (PDD)	$3 \times 3 \times 3$ , $b = 1500$	A Runge-Kutta 4 integration scheme was used.	[Tensaouti et al., 2009, Basser et al., 2000]	Sysiphe	Fatima Tensaouti
4	Single-DT	Tensor deflection: propagation direction follows direction resulting of the product between the current tensor and the previous direction	$3 \times 3 \times 3$ , $b = 1500$	Tensors are powered by a factor of 4 to increase anisotropy. A Euler integration scheme was used.	[Tensaouti et al., 2009, Lazar et al., 2003]	Sysiphe	Fatima Tensaouti
5	2-DT	Streamline tractography with filtered estimation of propagation direction	$3 \times 3 \times 3$ , $b = 1500$	The diffusion model estimation is guided by the previous propagation direction using unscented Kalman filtering.	[Malcolm et al., 2009, Malcolm et al., 2010]	3D Slicer	James Malcolm
6	PAS-MRI	Streamline with propagation direction following the PAS peak closest to previous direction (PAS amplitude was calculated on 181 points over the sphere)	$3 \times 3 \times 3$ , $b = 1500$	PAS was calculated by a fast algorithm [Sakaie, 2009]. Tracking is performed with the Fact algorithm of [Mori et al., 1999a] implemented in [Wang and Wedeen, 2007] from all phantom voxels and only fibers going through seeds were kept. For each seed, the longest fiber was chosen as the candidate fiber.	[Sakaie, 2009, Mori et al., 1999a, Wang and Wedeen, 2007]	Camino and TrackVis	Ken Sakaie
7	MoG	Global tractography	$3 \times 3 \times 3$ , $b = 2000$	Every point and direction of every fiber is a parameter of the model and contribute as a single isotropic Gaussian model.	[Reisert et al., 2009]	Matlab package	Marco Reisert
8	Single-DT	Streamline with propagation direction following tensor PDD	$3 \times 3 \times 3$ , $b = 1500$	Tracking is performed from all phantom voxels and only fibers going through seeds were kept. Spatial averaging of fibers was performed to return a single fiber / seed.	[Gouttard et al., 2009, Fillard et al., 2003]	MedNIRIA and Fiber-Tracking	Sylvain Gouttard
9	FOD-SH	Streamline with propagation direction following the FOD peak closest to previous direction	$3 \times 3 \times 3$ , $b = 1500$	Tracking is performed from all phantom voxels and only fibers going seeds were kept. A scoring function determines the most likely fiber to return a single fiber / seed.	[Yo et al., 2009, Descoteaux et al., 2009]	Brainvisa and Camino	Ting Yo
10	ODF-SH	Streamline with propagation direction following the FOD peak closest to previous direction	$3 \times 3 \times 3$ , $b = 1500$	A positivity constraint and spatial regularity are used during ODF estimation to make it more robust to noise.	[Goh, 2009, Goh et al., 2009a]	Matlab package	Alvina Goh

Rules of the contest impose to return a single fiber per voxel. Some participants had to post-process their tractography results that produce in general multiple fibers per voxel. Interesting post-processing was used to this end, from selection of the longest fiber to a more elaborate clustering method. While this was not expected, post-processing appears as necessary to improve the raw result of tractography algorithm.

In the next section, we present a qualitative review of the 10 contributions.

## 3.2 Qualitative Evaluation

We present on Figure 5 an overlap of the 10 contributions for each ground truth fiber (one image corresponding to one seed location), and on Figure 6 the individual results for each contribution (one image corresponds to the 16 candidate fibers of one method). Note that the image number does correspond to the method Id of Table 1. The first striking finding concerns the inter-method variability, which is relatively high depending on the seed location. It raises the fact that choosing one method or another can lead to completely different tracking results, and strengthen the importance of such comparative study to determine the optimal possibilities.

From Figure 5, we can conclude that, except for S13 and S14 that are located on the isolated U-shape structure (Region 7, Fig. 1 right), at least one contribution per seed fails at reconstructing the correct pathway. Very often the algorithm chose the wrong direction when going through crossing regions).

Seeds where only one contribution was mislead are S1, S3, S4 and S8. All contributions seem remarkably stable with those seeds. The most successful reconstructed fibers include those passing through S1, S3, and S4. Those were somehow the less difficult to reconstruct as the exact pathway traverses a single crossing region. Note that S2 also contains a single crossing but was less successful due to the fact that the seed was purposely located at the boundary of the phantom, thus leading to spurious partial voluming effect, specially when reaching crossing area 1 (Fig. 1 right) where three contributions chose the wrong branch. S8 was also very successful, as the seed was close to the boundary and tracking algorithms were able to correctly follow the large U-shape structure, successfully passing branching 5 and 6 (Fig. 1 right).

Among the less successful reconstructed fibers, we denote S5, S6, S7, S11, S12, S15 and S16: those had at least two crossing regions to traverse and many contributions were confused by one or the other.

Looking at the individual results presented on Figure 6 permits a more comprehensive inspection of the performance of each contribution. Visually, methods 2 and 7 seem to achieve the best reconstructions (we refer the reader to Table 1 for the Id - method correspondence). Method 10 is performing well, while fibers look very tortuous. In such situation, we expect the spatial metric to give a good score but the tangent and curve metrics to point this phenomenon out. Methods 1, 5, 6, and 9, although missing a few crossings, are performing reasonably well. Method 8 has fiber disruption in the central crossing area which is not a desirable feature. Besides crossings, most fibers appear as nicely reconstructed. Finally, methods 3 and 4 appear to be mislead in almost all crossings, although exhibiting smoother,

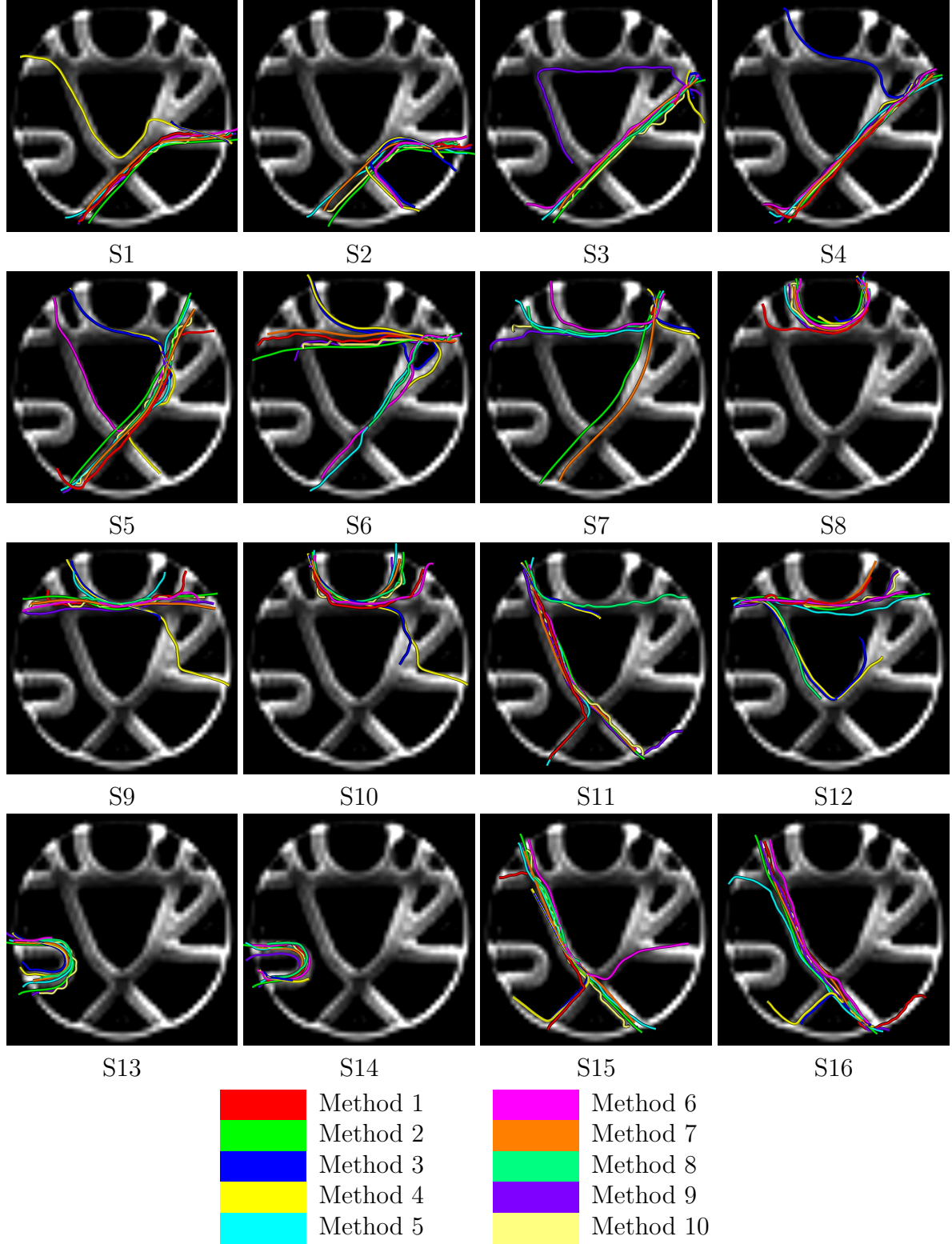


Figure 5: **Panel of images showing the reconstructed fiber of all contributions passing through each seed selected for the context.** The name below the image indicates the seed fibers originate (a color-method correspondence table is given below). Such overview allows to understand the variability of the results: some seeds (such as S3 and S4) were more successful than others (such as S7) to be reconstructed.

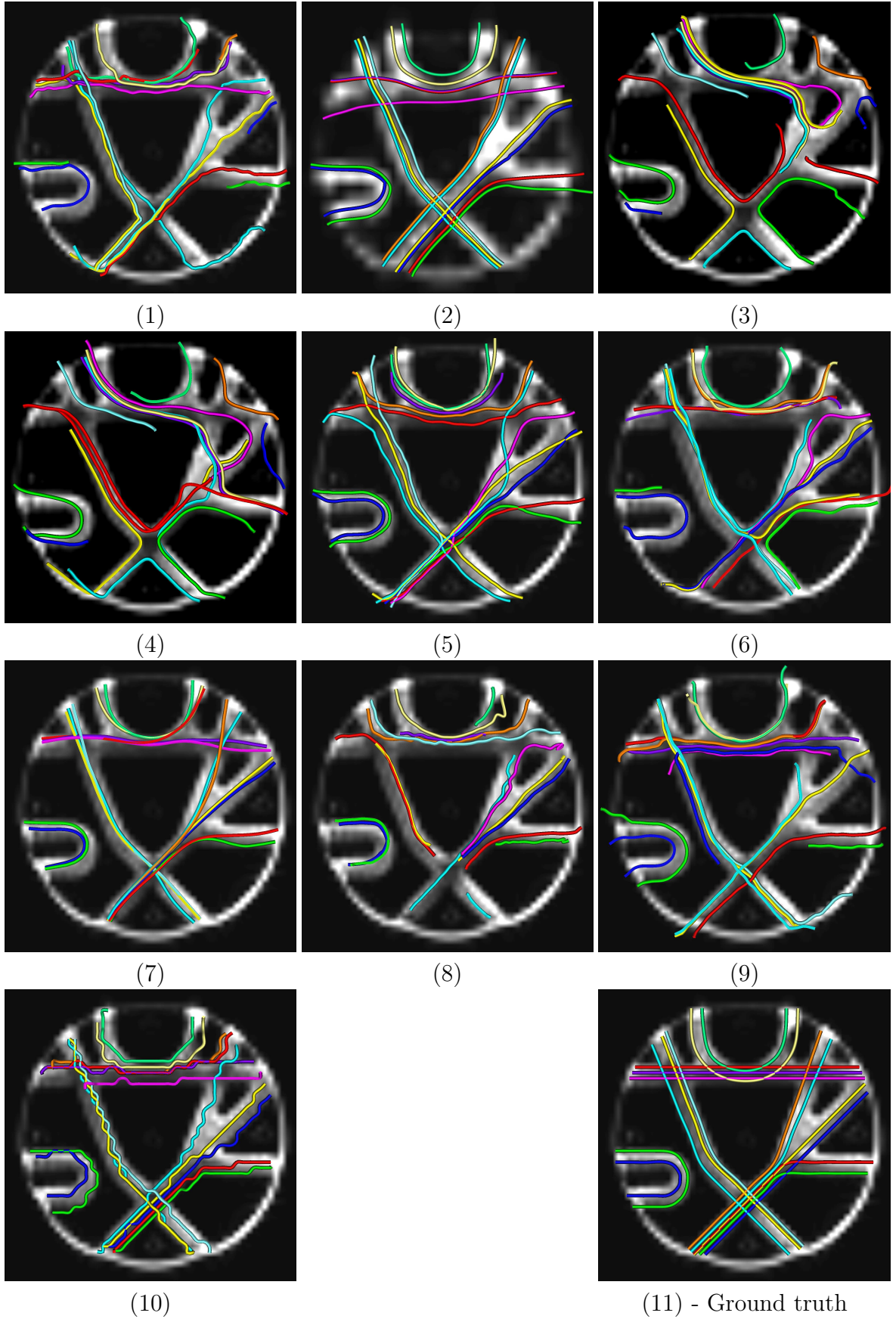


Figure 6: **Each of the 10 contributions is shown individually.** Image (11) represents the ground truth for visual comparison. Fibers are colored by the seed they originate (see Fig. 4 for more information on the seeds location). Visually, methods 2 and 7 seem to perform the best. We refer the reader to Table 1 for the method - Id correspondence.

and consequently more realistic, fibers than most of the other methods.

In the following, we present the quantitative evaluation of each of the 10 contributions.

### 3.3 Quantitative Evaluation

The symmetric RMSE of section 2.3 was evaluated for each of the three metrics and each of the 16 fibers of each method, which makes a total of:  $3 \times 16 \times 10 = 480$  sRMSE evaluated. One sRMSE comprises the evaluation of a metric on two times 1000 points, we end up with a total of 960 000 point-to-point metrics being tested. Results for the spatial, tangent and curve metrics are given on Figure 7.

The quantitative evaluation confirms our qualitative review of the results. Figure 7 presents results of the quantitative evaluation for the spatial, tangent and curve metrics. The colder the color in the metric figures, the better the performance of the algorithm. For the three metrics, methods 2 and 7 seem to outperform the others. For the spatial metrics, methods 1, 5, 9 and 10 perform well: fibers are indeed spatially very close to the ground truth, but some high values were detected for fibers where the tracking was misled in a crossing region. This is the case, for instance, for S7 where all methods except M2 and M7 have chosen the wrong pathway (see Fig. 5 (7)).

The tangent metric evaluates whether the fiber trajectory correctly follows the ground truth. Parallel fibers obtain a very good score for this metric. By contrast, fibers that chose the wrong pathway within a crossing will be highly penalized since the trajectory is very likely to become orthogonal to the correct one. Results presented on Figure 7 middle confirms this: method 2 realizes an almost perfect score since this is the only method which was able to correctly reconstruct all crossings. Method 7 got only misled in S12 as shown on Fig. 6 (7). Then, methods 5, 8 and 9 also exhibit perfect reconstructions for about half of the seeds. Methods 1, 3, 4 and 10 which produced noisier fibers than the others were penalized by this metric.

The curve metric of Figure 7 right penalizes fibers with high curvature since the ground truth contains only straight - or low-curved - fibers. It is a good indicator of the veracity of the result since fibers with high curvature are very unlikely to appear in real situations. Moreover, during phantom elaboration synthetic fibers were strongly stretched preventing high curvature to happen. Methods 2, 4, 5, 6 and 7 produced the fibers with the lowest curvature. Note that method 2 obtains bad scores for S3 and S14. Indeed, from Fig. 6 (3), one can notice that S3 has a non-smooth trajectory around branching 2 (Fig. 1 right) that appears like an inflection point. S14 has a maximum curvature at the far-right point of the U-shape structure (region 7, Fig. 1 right), which has been detected by the metric. Other methods, in particular method 10, are penalized by this metric due to the high frequency noise of their fibers. Note that the curve metric is not redundant with the tangent metric. For instance, method 4 received bad results for the tangent metric since fibers were obviously wrong, but good results for the curve metric since fibers were as smooth as the ground truth.

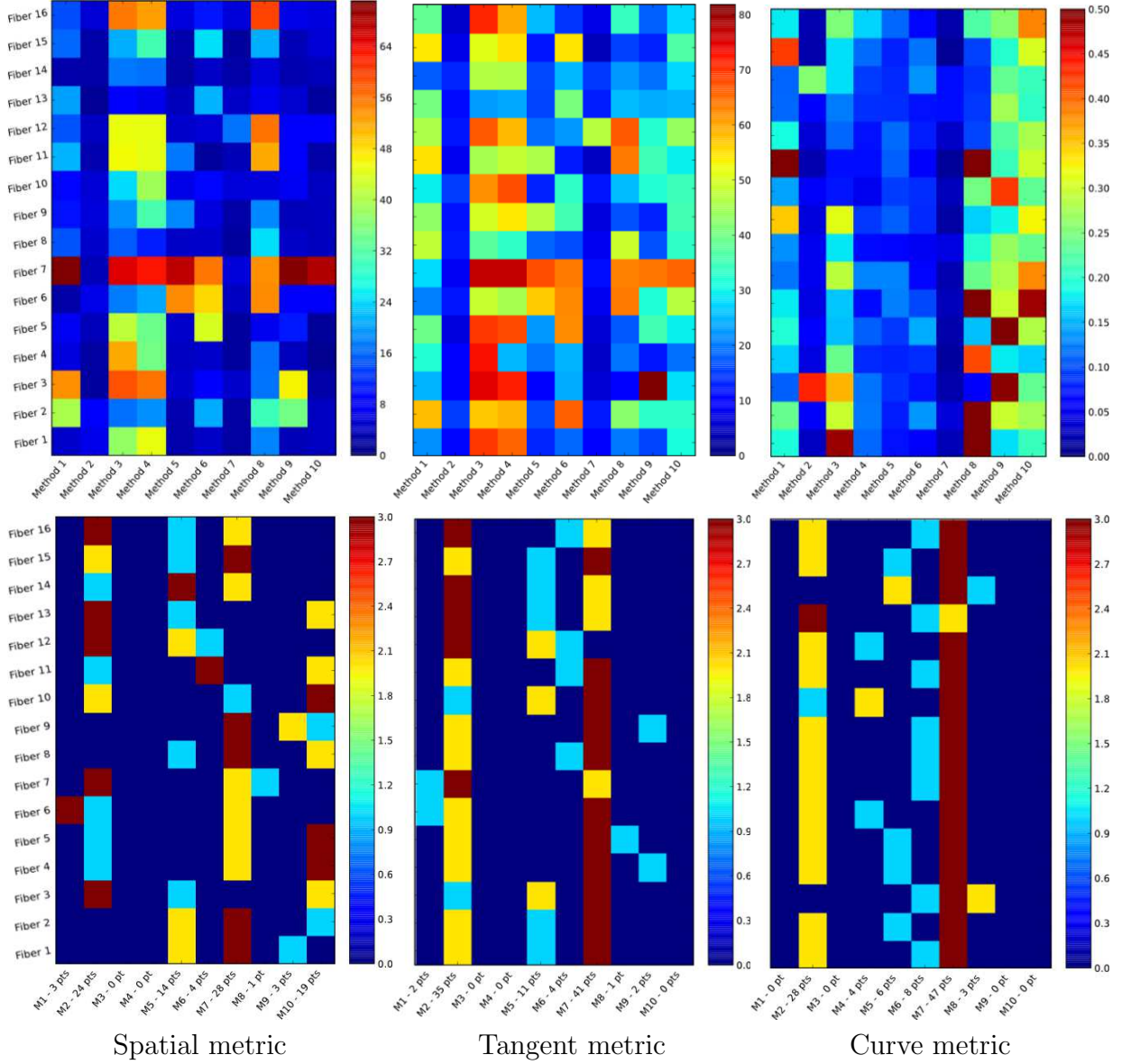


Figure 7: **Quantitative comparison of the 10 contributions to the ground truth using the spatial metric of Eq. 3 (left), the tangent metric of Eq. 4 (middle), and the curve metric of Eq. 5 (right). Top:** The metric scores for each method and each fiber. Units are in mm for the spatial metric, in degree for the tangent metric and in  $mm^{-1}$  for the curve metric. **Bottom:** Points attributed to each method. Color-coding is the following: dark blue: 0 point, light blue: 1 point, yellow: 2 points, and brown: 3 points. The total number of points for each method and each metric is indicated below the x-axis.

### 3.4 Ranking

Tractography results were ranked according to the following rule. For each fiber and each metric, the method realizing the best score (i.e., the lowest metric value) was attributed 3 points. The second method received 2 points, and the third 1 point. The other methods obtained 0 point. The purpose of such point attribution



is to give to each contribution a final and unique score to rank them. Obviously, improvements are possible since one may not desire to give the same importance to all metrics. However, this has the advantage to allow a simple comparison of the performance of different methods. Table 2 presents the final ranking of the submissions received for the contest. To further illustrate the performance of the tested algorithms in real situations, we performed tractography on a brain dataset with the top two methods (methods 7 and 2) and compared them to a single-DT streamline tractography algorithm (method 4). Results are presented in Supplementary Section 3.

Rank	Method	Score
1 <sup>st</sup>	<b>Method 7:</b> Global tractography	116 pts
2 <sup>nd</sup>	<b>Method 2:</b> FOD-SH with constrained spherical deconvolution and streamline tractography	87 pts
3 <sup>rd</sup>	<b>Method 5:</b> Combined 2-DT model estimation and streamline tractography	31 pts
4 <sup>th</sup>	<b>Method 10:</b> ODF-SH with positivity and regularity constraints and streamline tractography	19 pts
5 <sup>th</sup>	<b>Method 6:</b> PAS-MRI and streamline tractography	16 pts
6 <sup>th</sup>	<b>Method 1:</b> Adaptive 1 or 2-DT model and streamline tractography	5 pts
6 <sup>th</sup>	<b>Method 8:</b> Single-DT and streamline tractography	5 pts
6 <sup>th</sup>	<b>Method 9:</b> FOD-SH with streamline tractography	5 pts
9 <sup>th</sup>	<b>Method 4:</b> Single-DT with tensor deflection	4 pts
10 <sup>th</sup>	<b>Method 3:</b> Single-DT with streamline tractography and RK4 integration	0 pt

Table 2: **Final score ranking of the 10 contributions.** See Table 1 for more information about each method.

## 4 Discussion

### 4.1 Comments on the Methods

As expected, single tensor-based methods (Fig. 5 (3), (4) and (8)) seem to perform worse than others in crossing regions for the obvious reason that a single tensor is unable to correctly characterize the two-fiber compartment specific of those regions. In particular, in the lower crossing area (region 1, Fig. 1 right) methods 3 [Basser et al., 2000] and 4 [Lazar et al., 2003] chose to avoid it by contouring it, while method 8 [Fillard et al., 2003] stopped the tracking, very probably because the crossing yield a fiber curvature greater than the maximum angular deviation

authorized.

Multi-tensor based approaches (Fig. 5 (1) and (5)) are clearly a big improvement compared to single-tensor methods. Multi-tensor models can be expressed as a weighted sum of single tensors with the sum of the weights equals to one, each weight being proportional to the contribution of each tensor within a given voxel. However, those two approaches greatly differ. Method 1 (Fig. 5 (1)) [Ramirez-Manzanares et al., 2007] uses a mixture of single- and 2-tensor models. The 2-tensor is dropped at the profit of the single one when it is either detected as a single tensor, or when one of the tensor has a much larger importance than the other (i.e., the weight of one tensor is larger than twice the weight of the second). Indeed, in voxels with a single fiber bundle contribution, the single diffusion tensor model is often more reliable at evaluating the fiber direction. Then, a streamline tractography is used where the next propagation direction is chosen as the closest to the previous direction among all available candidate directions given the single or 2-tensor model.

Method 5 (Fig. 5 (5)) [Malcolm et al., 2010] elegantly combines the estimation of a 2-tensor model while performing tractography. Starting from a seed point, each fiber is traced by following the tensor whose PDD is the closest to the previous direction. However, instead of using least-squares to fit the tensor parameters directly, this method uses filtered estimation given the results of previous positions along the fiber. Specifically, it employs an unscented Kalman filter to provide a robust estimate of the highly nonlinear 2-tensor model, and of the mean and covariance of both tensor parameters. This produces a causal estimate of the local structure at each point along the fiber. Nevertheless, the streamline tractography algorithm used is the bottleneck of the method as errors may accumulate during the reconstruction, which may eventually lead to erroneous pathways.

A fast persistent angular structure calculation [Sakaie, 2009] was used in conjunction with a simple FACT tractography algorithm for method 6 (Fig. 5 (6)) on the 3 mm dataset with a b-value of 1500. Fibers were smooth, as indicated by good results of the curve metric, but some crossings were missed (regions 1, 4 and 5 Fig. of 1 right) leading to high values for the spatial and tangent metrics. Results were not improved when using the more extensive persistent angular structure calculation as implemented by Camino [Cook et al., 2006]. We can conclude that the high angular resolution and noise immunity of the persistent angular structure are not sufficient to compensate for shortcomings of simple streamline tractography in the presence of complex fiber geometries.

ODF- (Fig. 5 (10)) and FOD- (ODF with spherical deconvolution) based (Fig. 5 (2) and (9)) methods qualitatively give a good match with the ground truth. Notably, method 10 (Fig. 5 (10)) [Goh et al., 2009a] produced a very irregular and tortuous result. This is very likely to be caused by curve averaging as in a former submission competitors returned several fibers per seed which was not compliant with our requirements (a revised submission was then resent). Method 10 estimates the ODF using a probability density constraint and a spatial regularity prior. The constraint enforces the ODF to be positive, while the spatial prior ensures the resulting field to be spatially smooth, and the method is consequently robust to noise. Streamline tractography was then performed with a simple first



order integration scheme and by detecting ODF maxima by thresholding over the sphere. In the revised submission, the final curve for each seed is obtained by first computing all the possible fibers obtained from the ODF fields estimated from each repetition as well as the mean Riemannian ODF field, as described in [Goh et al., 2009b], and then discretizing the path that contains the most number of likely fibers. Such a tractography technique is noise-sensitive and at the same time, highly dependent on accurate ODF estimation. In spite of this limitation, the fact that only one fiber, going through S6, failed to cross region 3 (see Fig. 1 right), proves that the ODF estimation succeeded.

Method 9 (Fig. 5 (9)) estimates the FOD using the spherical deconvolution of [Descoteaux et al., 2009] with the constrained regularization of [Tournier et al., 2007]. While this method gives an acceptable qualitative match with the ground truth, it failed in reconstructing crossing region 4 (Fig. 1 right), and missed a couple of fibers in crossing region 1. Apparently, this method is noise-sensitive and the streamline tracking used is misled by erroneous FOD maxima, especially in crossing region 4 that exhibits a crossing with a low angular difference. Those regions requires a diffusion model with a good angular resolution. At the level of SNR of the chosen dataset, the ODF with spherical deconvolution was unable to correctly model this configuration.

Method 2 (Fig. 5 (2)) [Jeurissen et al., 2009, Jeurissen et al., 2010] presents very smooth fibers. This is also the only method which chose to use the  $6 \times 6 \times 6$  mm dataset instead of the  $3 \times 3 \times 3$  mm version. Apparently, the gain of SNR turned into a larger fiber regularity. Note also that all fibers take the correct pathway (no error in crossing or bending regions). Method 2 implements the constrained spherical deconvolution of [Tournier et al., 2007], giving an estimate of the FOD. Tractography was then conducted using a simple streamline approach based on an extraction of the FOD maxima using a Newton optimization, which tends to prove, given the quality of the results, that the combination of spatially-adaptive DWI denoising before sharp FOD estimation and maxima extraction is a good choice of processing steps. Indeed, the gain in SNR with the 6 mm dataset, further increased with the denoising, seem to overcome the decreased precision of the fiber spatial positions induced by the diminished resolution. We may argue that SNR plays a key role in diffusion model estimation and should not always be sacrificed at the profit of spatial resolution.

Method 7 (Fig. 5 (7)) [Reisert et al., 2009], which used global tractography, realized the best scores for all metrics and was declared the winner of the contest. Global tractography is a class of tractography algorithms that can be seen as greedy algorithms: instead of reconstructing tracts one by one, those algorithms reconstruct all fibers at the same time [Mangin et al., 2002, Kreher et al., 2008, Fillard et al., 2009]. More precisely, in global tractography each segment of a fiber (i.e., an oriented point) is a parameter to be optimized. Optimization is performed such that each segment tries to associate with neighboring segments to form longer chains of low curvature while modelling the diffusion weighted data at best. Each fiber segment contributes as a single isotropic Gaussian model, which eventually results in a mixture of Gaussian in each voxel. Those methods are a lot more time-consuming than their deterministic counterparts but seem to be well-adapted in real, noisy situations.

## 4.2 Recommendations

Although this evaluation is subject to some limitations (which are discussed in the next section), it is still possible to make a few recommendations about methods which should be used and those which should be avoided in tractography. The recommendations that follow are based on the tested implementations of each method. Other implementations could have conducted to different conclusions. Moreover, there is no guarantee that the results obtained on the phantom dataset can be directly transposed to real situations. While MR phantoms are undoubtedly a step towards the validation of diffusion MRI tractography, they should be complemented by other sources of known fiber pathways such as post-mortem data.

First, in the context of tractography, the single-tensor model should be used with extra-caution since tractography algorithms are very likely to fail in reconstructing crossing fibers. However, the single-DT model is still able to correctly characterize numerous fiber bundles as shown in the real experiment of Supplementary Section 3. Notably, the DT model with only few degrees of freedom is by essence less sensitive to noise than more complex models, which often makes it the unique alternative in clinical applications.

Second, in case of good quality datasets, the best option seems to use a fiber orientation distribution function in conjunction with a streamline tractography algorithm where the next direction of propagation is directly inferred from the FOD maxima. Indeed, with reasonable SNR datasets, FODs seem successful in modelling the fiber directions within a voxel and can be trusted.

Finally, for datasets of medium and low quality as it is often encountered in real situations, several options are possible but all of them are using a spatial prior to make the model estimation more robust to noise. For instance, method 5 constraints the estimation of a 2-tensor model by the previous fiber direction, which gives some spatial regularity to the whole procedure. Method 10 explicitly imposes a spatial regularity when estimating the ODF, which eventually leads to good fiber pathways even using a streamline tractography algorithm, which give some evidence that the fiber directions were correctly modeled by the ODF. The global tractography of method 7 also uses a smoothness prior during fiber reconstruction (fiber segments associate to form chains of low curvature). Additional spatially-adaptive smoothing of the DWI is also beneficial, as shown by method 2. Conversely, without spatial prior, not any diffusion model was shown to correctly estimate the different fiber contributions within a voxel, and consequently should be used with extreme caution.

## 4.3 Limitations and Future Work

One of the caveat of the proposed quantitative evaluation methodology is that it only applies to deterministic tractography algorithms. Outputs of probabilistic methods are maps of connectivity that give, at every voxel of a regular 3D grid, the probability to be connected to a reference position. A ground truth is more difficult to establish as intrinsic factors such as fiber density inhomogeneities, fiber diameter or tissue permeability may impact the motion of water molecules. Thus, elaboration of a ground truth dataset for probabilistic tractography requires careful investigations and was out of scope of the present work. The purpose of this work is to give evidence that reconstructed pathways are correctly representing the true fiber organization, which is crucial for many applications like neurosurgery planning.

Furthermore, the synthetic phantom does not properly simulate the physiology of the diffusion phenomenon as it happens within living tissues. Indeed, only extra-cellular diffusion is simulated since no water molecules diffuse within synthetic fibers (diffusion only occurs around). A side effect is that the values of the fractional anisotropy within the fiber bundles are lower than those generally observed in brain white matter, which tends to disadvantage tensor-based methods. However, methods based on diffusion models like ODF with spherical deconvolution appears to be less sensitive to this diminished anisotropy and perform well at separating the various bundle contributions, making the overall evaluation method still valid.

In addition, we would like to recall that the results presented here rely on a blind contest: competitors did not have access to the ground truth and were not authorized to modify their results after the ground truth was revealed. Consequently, results may be sub-optimal, as fine parameter tuning could have improved tractography accuracy. Now that the ground truth is revealed, it is possible that the ranking of the methods would be different if the analysis were to be repeated. Furthermore, this study was performed on a single dataset. Therefore, the recommendations we make are purely indications and should not be considered as established truth. It is only by repeating this type of analysis on different datasets, and by analyzing the consistency of the results, that we will be able to conclude about the superiority of one model/tractography algorithm compared to another.

In the future, new evaluation criteria will be proposed. Notably, some characteristics of the phantom were left unexploited. For instance, it is possible to evaluate the angular differences between crossing bundles and compare these values to the ground truth (see Fig. 1). Another possibility is to evaluate whether the boundaries of a bundle are correctly reconstructed by measuring the spatial distance in-between two tracts delimiting the bundle. Finally, adding new seeds to the evaluation would allow a rapid extension of the current work.

To conclude, we want to emphasize that, to the best of our knowledge, this is the first quantitative evaluation of multiple tractography methods on a common dataset with known ground truth. We believe that such a common dataset along with the methodology proposed here can serve as an evaluation basis for existing and new algorithms. To this end, the DWI, the ground truth fibers, the evaluation methodology and the results obtained so far will remain freely available on a permanent website: <http://www.lnao.fr/spip.php?rubrique79>. New results can be submitted for evaluation by emailing them to [fibercup09@gmail.com](mailto:fibercup09@gmail.com). Results will be ranked and published among others on the *Fiber Cup* website.

## Acknowledgements

The mechanical conception of the phantom was supervised by Cyril Poupon, Laurent Larivière, Grégory Tournier, and Denis Fournier. Its manufacturing was taken in charge by Jérémy Bernard, Grégory Tournier and Laurent Larivière from the mechanics lab of the NeuroSpin centre, under the supervision of Laurent Larivière. The MRI acquisitions were realized under the supervision of Cyril Poupon and Irina Kezele. The organizers would like to thank Dr Ghislaine Dehaene (LCOGN Lab,

NeuroSpin centre) for funding this project.

## 5 Supplementary Material

### Supplementary section 1: Probabilistic Tractography of the Phantom Dataset using FSL

In this section, we demonstrate the ability of the probabilistic tractography algorithm of FSL [Behrens et al., 2007a] to produce connectivity maps (CM) characterizing the various bundles contained in the phantom dataset. To remain consistent with the choice of the competitors, the algorithm was run on the 3 mm dataset with a b-value of 650. Results are presented on Figure 8 and discussed below.

The qualitative evaluation of the CM was done by considering whether all voxels of the bundle a seed belongs to were selected by the algorithm without including voxels of *a priori* distinct bundles. Based on this setting, the following comments were inferred. The method succeeds in seeds S3, S4, S13, S14 and also S8 as the U-shape structure appears well connected to the seed. CM from S1 presents an undesired branching in region 1, although the pathway with the highest connectivity index remains correct. CM from seeds S9, S11 and S15 appear to “leak” due to crossings in region 1, 3 and 4 but still connect all voxels of the targeted bundles. CM from seed S5, however, could not pass region 4, which is a limitation. CM from seeds S6 and S7 were misled by the crossing in region 4 similarly to most of the deterministic algorithms evaluated. Finally, S2, S12 and S16 ended prematurely due to crossings in regions 1 and 2.

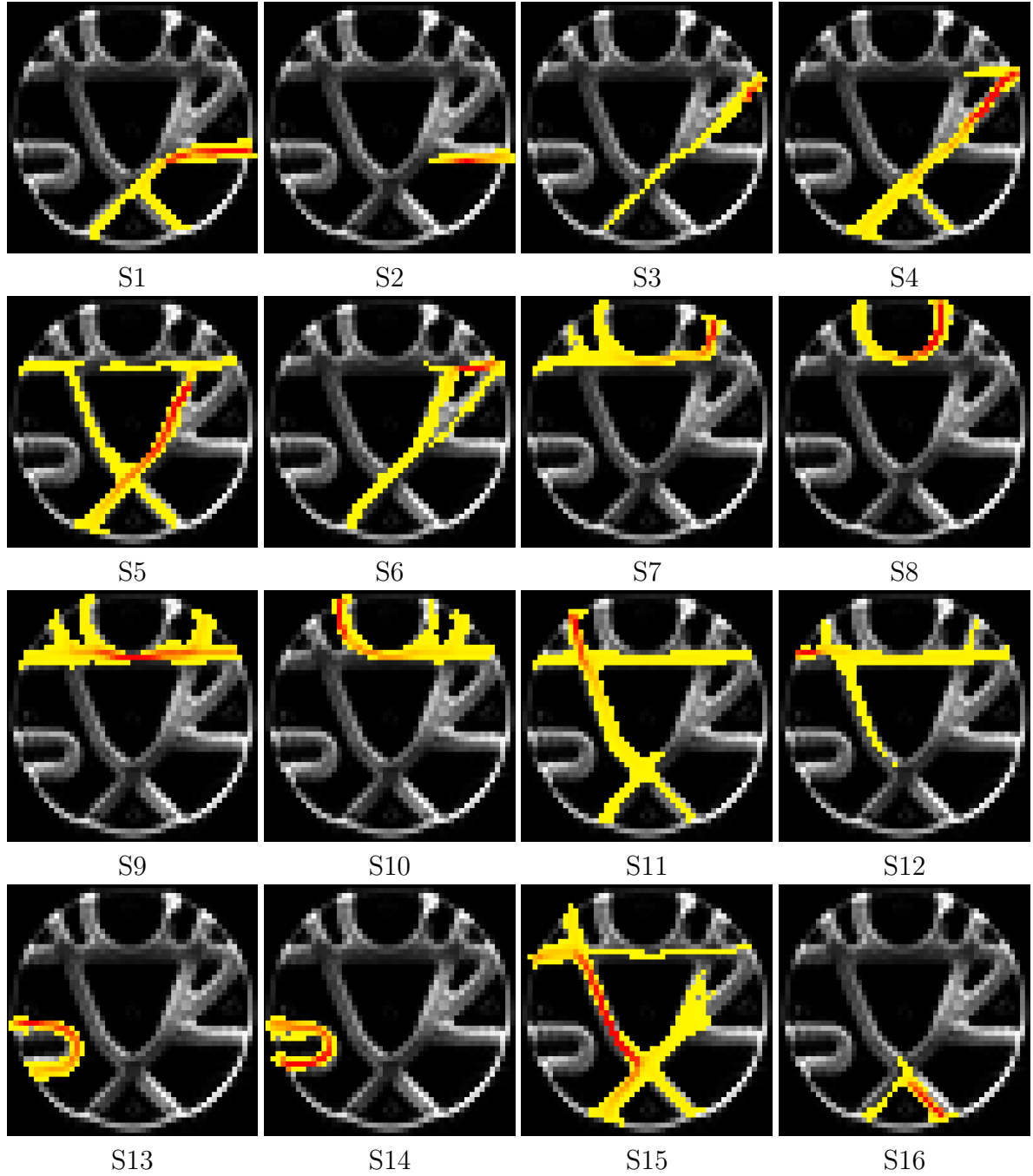


Figure 8: **Probabilistic tractography of the phantom dataset using FSL.** The connectivity maps are shown on top of the  $b_0$  image. We refer the reader to Figure 4 for the seed locations. The color range goes from yellow (low probability of connexion) to red (high probability). The obtained CMs indicate that probabilistic tractography remains feasible on such synthetic dataset, which could be further used to evaluate this class of algorithm.

## Supplementary section 2: Details for Result Reproducibility

In this section, the necessary details for the reproducibility of the tractography results of the 10 methods are given. For each of them, contact information are given.

### Method 1: Adaptive 1- or 2-DT and Streamline Tractography

**Contact:** Alonso Ramirez-Manzanares (aram@cimat.mx).

A MATLAB implementation of the method is available online along with a web page listing instructions to perform tractography on the phantom dataset: <http://www.cimat.mx/~aram/fibreCup/>.

#### Procedure:

- Download the MATLAB files: [http://www.cimat.mx/~aram/fibreCup/dbf\\_tracto\\_fibreCupData.zip](http://www.cimat.mx/~aram/fibreCup/dbf_tracto_fibreCupData.zip).
- Add the install path to the MATLAB path environment variable.
- Edit the MATLAB script file DBF\_matlab/main\_launcher.m, and replace the occurrence of installation\_folder with the path where the program was installed.
- Run the script DBF\_matlab/main\_launcher.m.

### Method 2: FOD with Constrained Spherical Deconvolution and Streamline Tractography

**Contact:** Ben Jeurissen (ben.jeurissen@ua.ac.be).

The tractography method has been incorporated into the EXPLOREDTI toolbox [Leemans et al., 2009], available at request on <http://www.exploredti.com>.

#### Procedure:

- Under Settings, Tractography, Algorithm: select the Streamline CSD algorithm.
- Under Settings, Tractography, Parameters: set the following parameters: Step size: 0.1 mm; FOD threshold: 0.1; angle threshold: 60 degrees.

Please note that this does not include the spatially-adaptive smoothing preprocessing step.

### Method 3: Single-DT with Streamline Tractography and Runge-Kutta 4 Integration

**Contact:** Fatima Tensaouti (tensaouti.f@chu.toulouse.fr).

Implementation of the method is provided in the SISYPHE software available on request.

**Procedure:**

- Set Fractional Anisotropy Threshold to 0.01.
- Set Maximum Angle of Deviation to 45 degrees.
- Launch RK4 Euler Tractography.

The step size for RK4 Euler integration is fixed to 1 mm.

**Method 4: Single-DT with Tensor Deflection Tractography**

**Contact:** Fatima Tensaouti (tensaouti.f@chu.toulouse.fr).

Implementation of the method is provided in the SISYPHE software available on request.

**Procedure:**

- Set Fractional Anisotropy Threshold to 0.01.
- Set Maximum Angle of Deviation to 45 degrees.
- Set Anisotropy Enhancing Exponent to 4.
- Launch Tensor Deflection Tractography.

The step size for Euler integration is fixed to 1 mm.

**Method 5: 2-DT with Filtered Tractography**

**Contact:** James Malcolm (malcolm@bwh.harvard.edu).

A SLICER implementation of the method is being developed but is not available at that time.

**Method 6: PAS-MRI with FACT Tractography**

**Contact:** Ken Salaie (sakaiek@ccf.org).

The software **Camino** was used for PAS-MRI evaluation. TRACKVIS was used for fiber tractography. Both were combined in a homemade software available under request.

**Procedure:**

- PAS was calculated at each voxel with a point set of size 1024. This was accomplished with CAMINO with the -fastmesd option.
- Amplitudes of the PAS function was calculated at 180 points as specified by TRACKVIS ([http://www.trackvis.org/dtk/181\\_vecs.txt](http://www.trackvis.org/dtk/181_vecs.txt)) and a binary data file was saved according to the format specified by TRACKVIS (<http://www.trackvis.org/dtk/?subset=format>).
- Tractography was run with the default options using the TRACKVIS command-line program, odf\_tracker: step size: 0.1 mm, angular threshold: 35 degrees, Runge-Kutta 2 integration scheme.



## Method 7: Global Tractography

**Contact:** Marco Reisert (marco.reisert@uniklinik-freiburg.de).

A Matlab implementation of global tractography is available under request. Fibers were calculated using the following procedure.

### Procedure:

- Extract a tracking mask by thresholding the  $b_0$  image at 200.
- Start the tracker using command `fiberGT_tool`.
- Load the hardi data via Load HARDI button.
- Load the mask via Load Mask button.
- Press the dense parameter button.
- Press start tracking.

## Method 8: Single-DT and Streamline Tractography

**Contact:** Sylvain Gouttard (gouttard@sci.utah.edu).

The MEDINRIA software version 1.8 was used for single-DT estimation. The FIBERTRACKING software was used for streamline tractography. The procedure is as follows.

### Procedure:

- Open MEDINRIA and import diffusion data into a new DTI study.
- Under the Tensor Processing tab, set the Background Removal Threshold to 200.
- Click on the Estimate Model button from the menu and save tensors in a vectorial file format (such as NIFTI).
- Open FIBERTRACKING and load the tensor field (File then Load Tensor Image).
- From the tab Parameters, set the White Matter Extraction Threshold to 0.02.
- From the Pre-Processing tab, click on Compute FA\_MD.
- As this tool takes only one seed point region at a time, the seed image provided was first split into 16 images with one one voxel at the value 1. From the Computation tab, load the seed image using the Load ROIs button. This has to be repeated for every seed.
- Select Track the whole brain check box.
- From the Parameters tab, set the Fiber Tracking Threshold and Angle of Max Deviation to the values listed in the Table 3 above. Values are different for each seed.
- From the Computation tab, press the Compute Fibers! button.

The step size for Euler integration is fixed to 1 mm.

Seed	Fiber Tracking Threshold	Max. Angle
1	0.780	0.102
2	0.804	0.102
3	0.715	0.102
4	0.897	0.102
5	0.808	0.102
6	0.724	0.102
7	0.668	0.102
8	0.701	0.102
9	0.937	0.102
10	0.741	0.102
11	0.506	0.500
12	0.693	0.500
13	0.693	0.102
14	0.813	0.898
15	0.909	0.102
16	0.709	0.102

Table 3: **Seed-dependent parameters for method 8.**

## Method 9: FOD using Spherical Harmonics and Streamline Tractography

**Contact:** Ting Yo (tingshuo.yo@gmail.com).

The BRAINVISA software package was used for ODF-SD estimation while CAMINO was used for tractography. The detailed procedure is as follows:

### Procedure:

- ODF-SD estimation was performed using the `PtkDwiOdfField` command available in BRAINVISA:  

```
PtkDwiOdfField -t2 <b0 image> -dw <dwis> -m <image mask> -type
sdt_odf_cartesian_field -f minmax_normalized_mesh -o mesho8 -f odf_sh
-o sho8_k2 -outputOrientationCount 1000 -verbose true -scalarParameters
0 2.05e-9 0.86 8 0.006 1
```
- Convert ODF-SD BRAINVISA format to CAMINO: `mesho8 -> b1500_sho8.Bfloat`. Conversion utilities are available on request (tingshuo.yo@gmail.com).
- Extract ODF-SD principal directions using the `sfpeaks` command of Camino:  

```
sfpeaks -inputfile b1500_sho8.Bfloat -inputdatatype float -inputmodel
sh -order 8 -numpds 3 -schemefile b1500.scheme -stdsfrommean 4 >
b1500_pds.Bdouble
```
- Prepare the anisotropy map: `b1500_fa.hdr`.
- Prepare the seed file by thresholding the anisotropy map at 0.05: `b1500_fa05.hdr`.
- Perform tractography with the `track` command of Camino:

```
track -inputmodel pds -numpds 3 -inputfile b1500_pds.Bdouble -seedfile
b1500_fa05.hdr -anisthresh 0.05 -anisfile b1500_fa.hdr -interpolate -stepsize
0.5 -curvethresh 60 -outputfile tracts/b1500_k2o8_fa05_c60.Bfloat
```

The step size for integration was 0.5 mm, and the angle threshold was set to 60 degrees.

## Method 10: ODF with Positivity Constraint and Spatial Regularity and Streamline Tractography

**Contact:** Alvina Goh (agoh@jhu.edu).

A Matlab implementation of FOD estimation with positivity constraint and spatial regularity is available under request. Fibers were calculated using the following procedure.

### Procedure:

- FOD are estimated using the `odfestimate` matlab function as follows:

```
odfestimate -input data -SHorder 4 -reconstructiontess 4 -output ODFSH.mat
```

- Fibers were tracked using the `fibertrack` matlab function:

```
fibertrack -seedposition [x,y,z] -eliminatemaxangle 45 -incominglocalanglethreshold
20 -possiblecurveanglethreshold 60 -possiblecurvemagthreshold 0.8 -nomaxmagthreshold
0.7 -output possiblefibertracts.mat
```

The step size for integration is 1 mm. The angle threshold was set to 60 degrees.

- Finally, fibers were further refined using the `calculatepathcrossedmaxtimes` matlab function:

```
calculatepathcrossedmaxtimes -input possiblefibertracts.mat -output final-
tracts.mat
```

### Supplementary section 3: Tractography Examples on a Brain Dataset

In order to illustrate the performance difference of the winning methods, we performed tractography on a real brain dataset with method 7 and 2 and compared them to a single-DT method (method 4). To present the results, two ROIs were manually defined and only fibers passing through the ROIs were retained. The first ROI delineates the corpus callosum (Figure 9 left): the bundle should not only contain the U-shaped cortico-cortical fibers, but also the longer projection fibers that pass several regions of crossing fibers in the brain (including the corona radiata). Results are shown on Figure 10 left. The second ROI delineates the pons (Figure 9 right). Fibers passing through the pons should go up to the motor cortex via the corticospinal tract and fan along the right and left hemisphere. Results are shown on Figure 10 right.

As expected, all methods were able to successfully reconstruct the U-shape callosal fibers (Figure 10 left). However, the single-DT method was not able to reconstruct the longer projection fibers that cross multiple other bundles such as the corona radiata, while the two other methods succeeded. The red color of the fibers indicate that they are mostly left-right oriented, which is coherent with what was expected.

Similarly, all methods succeeded in reconstructing fibers passing through the pons and going up to the motor cortex (Figure 10 right). However, only Method 7 (and partly method 2) was able to detect the fanning structure of the cortico-spinal tract.

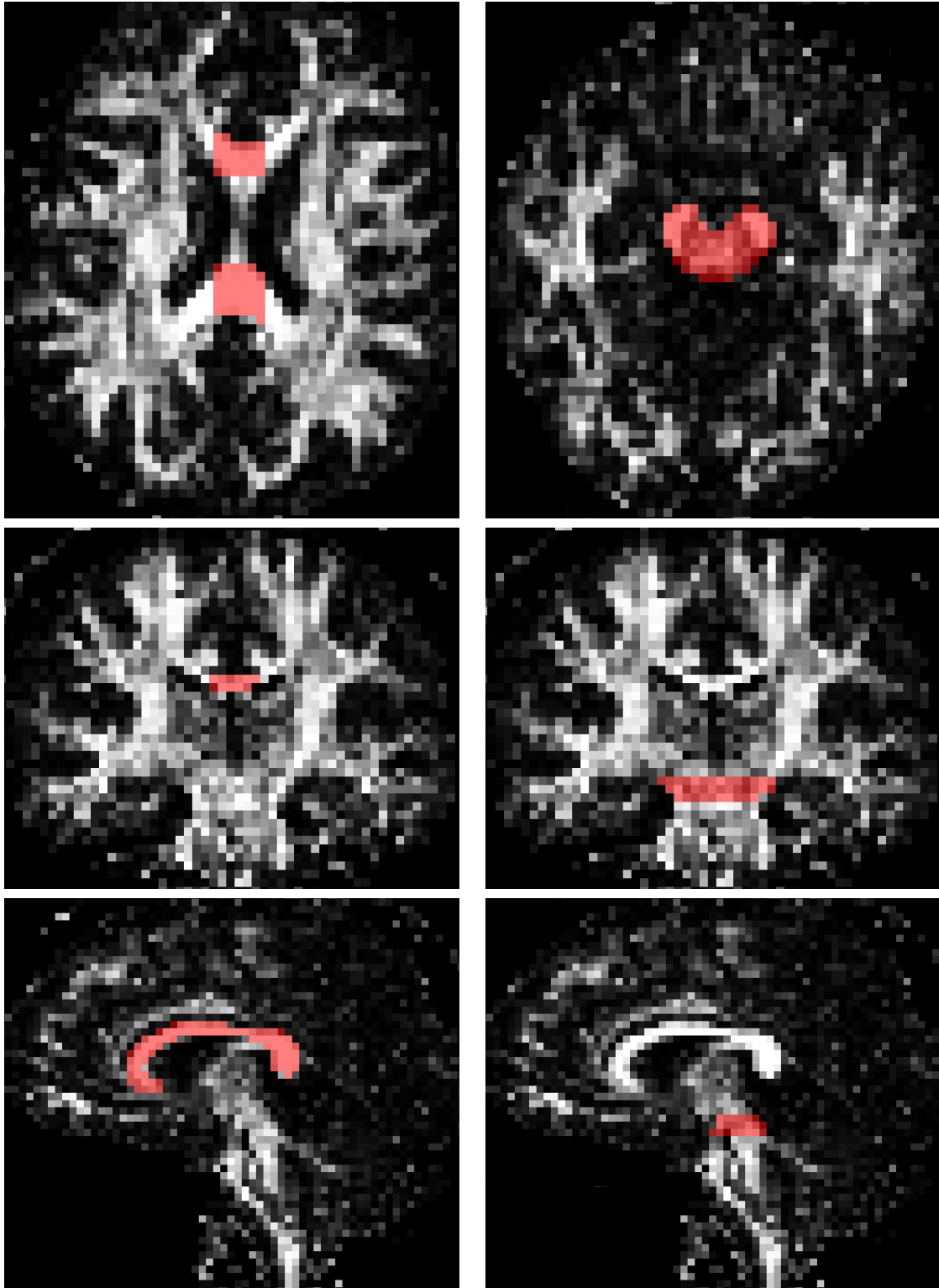
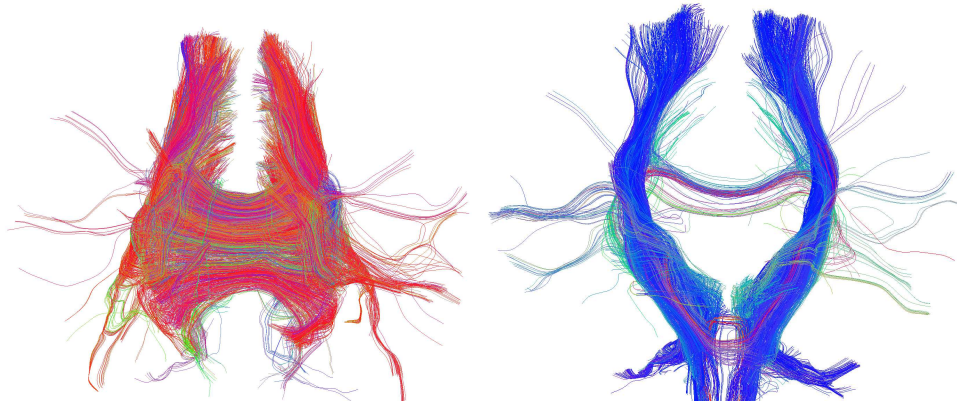
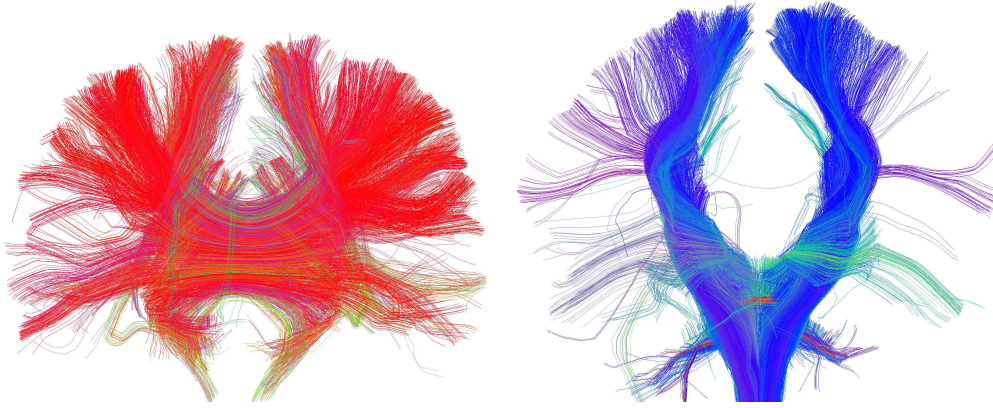


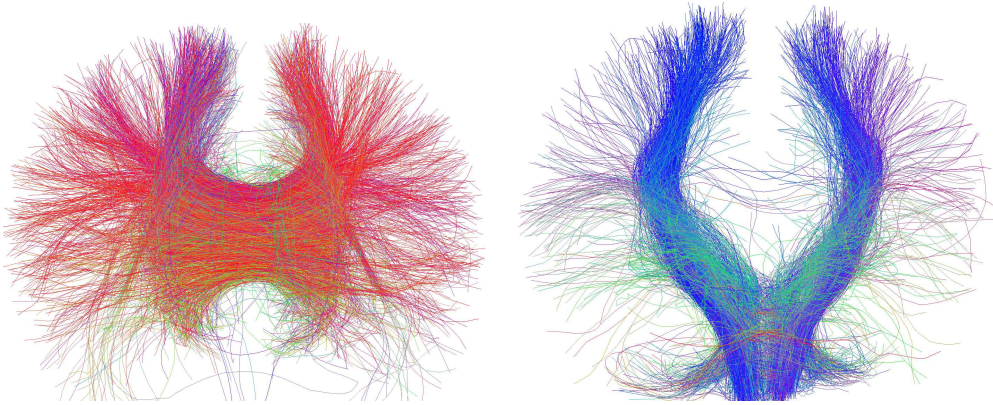
Figure 9: **ROI used for tractography on a real dataset. Left:** Corpus Callosum. **Right:** Pons. **Top:** Axial view. **Middle:** Coronal view. **Bottom:** Sagittal view. ROI were manually drawn on the FA map.



Single-DT and Streamline Tractography (method 4)



ODF with Constrained Spherical Deconvolution and Streamline Tractography (method 2)



Global Tractography (method 7)

Figure 10: **Examples of tractography on a real dataset of three methods.** For each method, the entire white matter was reconstructed and only fibers passing through the ROIs shown on Figure 9 are shown. **Left:** Fibers passing through the corpus callosum (Fig. 9 left). **Right:** Fibers passing through the pons (Fig. 9 right). Results are shown for: **top:** single-DT and streamline tractography (method 4), **middle:** FOD with constrained spherical deconvolution and streamline tractography (2nd best method), and **bottom:** for global tractography (winning method). The color codes the main fiber orientation: a left-right oriented fiber is colored in red, an infero-superior fiber is blue and an antero-posterior fiber is green.

## References

- [Anderson et al., 2006] Anderson, A., Choe, A., Stepniewska, I., and Colvin, D. (2006). Comparison of brain white matter fiber orientation measurements based on diffusion tensor imaging and light microscopy. In *Proceedings of the IEEE Engineering in Medicine and Biology Society*, pages 2249–51.
- [Assaf and Basser, 2005] Assaf, Y. and Basser, P. (2005). Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. *NeuroImage*, 27(1):48–58.
- [Basser et al., 1994] Basser, P., Mattiello, J., and LeBihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *Journal of Magnetic Resonance B*, 103(3):247–254.
- [Basser et al., 2000] Basser, P., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magnetic Resonance in Medicine*, 44:625–632.
- [Behrens et al., 2007a] Behrens, T., Johansen-Berg, H., Jbabdi, S., Rushworth, M., and Woolrich, M. (2007a). Probabilistic diffusion tractography with multiple fibre orientations. what can we gain? *NeuroImage*, 23:144–155.
- [Behrens et al., 2007b] Behrens, T. E. J., Johansen-Berg, H., Jbabdi, S., Rushworth, M. F. S., and Woolrich, M. W. (2007b). Probabilistic diffusion tractography with multiple fibre orientations. what can we gain? *NeuroImage*, 34(1):144–155.
- [Behrens et al., 2003] Behrens, T. E. J., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., Matthews, P. M., Brady, J. M., and Smith, S. M. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magnetic Resonance in Medicine*, 50:1077–1088.
- [Bergmann et al., 2007] Bergmann, Ø., Kindlmann, G., Peled, S., and Westin, C.-F. (2007). Two-tensor fiber tractography. In *4th International Symposium on Biomedical Imaging*, pages 796–799, Arlington, Virginia, USA.
- [Campbell et al., 2006] Campbell, J. S. W., Savadjiev, P., Siddiqi, K., and Pike, G. B. (2006). Validation and regularization in diffusion MRI tractography. In *Third IEEE International Symposium on Biomedical Imaging (ISBI): from Nano to Macro*, pages 351–354, Arlington, Virginia, USA.
- [Chao et al., 2007a] Chao, Y.-P., Yang, C.-Y., Cho, K.-H., Yeh, C.-H., Chou, K.-H., Chen, J.-H., and Lin, C.-P. (2007a). Probabilistic anatomical connection derived from QBI with MFACT approach. In *International Conference on Functional Biomedical Imaging*, Hangzhou, China.
- [Chao et al., 2007b] Chao, Y.-P., Yeh, C.-H., Cho, K.-H., Chen, J.-H., and Lin, C.-P. (2007b). Multiple streamline tractography approach with high angular resolution diffusion imaging data. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 1550, Berlin, Germany.
- [Ciccarelli et al., 2008] Ciccarelli, O., Catani, M., Johansen-Berg, H., Clark, C., and Thompson, A. (2008). Diffusion-based tractography in neurological disorders: concepts, applications, and future developments. *The Lancet Neurology*, 7(8):715–27.

- [Conturo et al., 1999] Conturo, T., Lori, N., Cull, T., Akbudak, E., Snyder, A., Shimony, J., McKinstry, R., Burton, H., and Raichle, M. (1999). Tracking neuronal fiber pathways in the living human brain. *Proceedings of the National Academy of Sciences*, 96:10422–10427.
- [Cook et al., 2006] Cook, P., Bai, Y., Nedjati-Gilani, S., Seunarine, K., Hall, M., Parker, G., and Alexander, D. (2006). Camino: Open-source diffusion-mri reconstruction and processing. In *Proceedings of the 14th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*.
- [Dauguet et al., 2007] Dauguet, J., Peled, S., Berezovskii, V., Delzescaux, T., Warfield, S. K., Born, R., and Westin, C.-F. (2007). Comparison of fiber tracts derived from in-vivo dti tractography with 3d histological neural tract tracer reconstruction on a macaque brain. *Neuroimage*, 37(2):530–538.
- [Descoteaux et al., 2009] Descoteaux, M., Deriche, R., Knosche, T., and Anwander, A. (2009). Deterministic and probabilistic tractography based on complex fibre orientation distributions. *IEEE Transactions on Medical Imaging*, 28(2):269–286.
- [Fieremans et al., 2008] Fieremans, E., Deene, Y., Delpitte, S., Özdemir, M., Achten, E., and Lemahieu, I. (2008). The design of anisotropic diffusion phantoms for the validation of diffusion weighted magnetic resonance imaging. *Physics in medicine and biology*, 53:5405.
- [Fillard et al., 2007] Fillard, P., Arsigny, V., Pennec, X., Hayashi, K. M., Thompson, P. M., and Ayache, N. (2007). Measuring brain variability by extrapolating sparse tensor fields measured on sulcal lines. *Neuroimage*, 34(2):639–650. Also as INRIA Research Report 5887, April 2006. PMID: 17113311.
- [Fillard et al., 2003] Fillard, P., Gilmore, J., Lin, W., Piven, J., and Gerig, G. (2003). Quantitative analysis of white matter fiber properties along geodesic paths. In *Proc of MICCAI’03, Part I*, volume 2879 of *LNCS*, pages 16–23. Springer.
- [Fillard et al., 2009] Fillard, P., Poupon, C., and Mangin, J.-F. (2009). A novel global tractography framework based on an adaptive spin glass model. In *Proc. 12th MICCAI*.
- [Friman et al., 2006] Friman, O., Farneback, G., and Westin, C.-F. (2006). A bayesian approach for stochastic white matter tractography. *IEEE Transactions in Medical Imaging*, 25(8):965–977.
- [Goh, 2009] Goh, A. (2009). Deterministic tractography using orientation distribution functions estimated with probability density constraints and spatial regularity. In *MICCAI workshop on Diffusion Modelling and the Fiber Cup (DMFC’09)*, London, United Kingdom.
- [Goh et al., 2009a] Goh, A., Lenglet, C., Thompson, P. M., and Vidal, R. (2009a). Estimating orientation distribution functions with probability density constraints and spatial regularity. In *In Proc. of MICCAI*, volume 12, pages 877–885.
- [Goh et al., 2009b] Goh, A., Lenglet, C., Thompson, P. M., and Vidal, R. (2009b). A nonparametric Riemannian framework for processing high angular resolution diffusion images (HARDI). In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2496–2503.



- [Gouttard et al., 2009] Gouttard, S., Sharma, A., and Gerig, G. (2009). Fiber challenge - sci utah solution. In *MICCAI workshop on Diffusion Modelling and the Fiber Cup (DMFC'09)*, London, United Kingdom.
- [Hagmann et al., 2004] Hagmann, P., Reese, T. G., Tseng, W.-Y. I., Meuli, R., Thiran, J.-P., and Wedeen, V. J. (2004). Diffusion spectrum imaging tractography in complex cerebral white matter: an investigation of the centrum semiovale. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 623. International Society for Magnetic Resonance in Medicine.
- [Hall and Alexander, 2009] Hall, M. and Alexander, D. (2009). Convergence and parameter choice for monte-carlo simulations of diffusion mri. *IEEE Transactions on Medical Imaging*, 28(9):1354–64.
- [Haroon and Parker, 2007] Haroon, H. A. and Parker, G. J. (2007). Using the wild bootstrap to quantify uncertainty in fibre orientations from q-ball analysis. In *Proceedings of the International Society of Magnetic Resonance in Medicine*, page 903, Berlin, Germany.
- [Jansons and Alexander, 2003] Jansons, K. M. and Alexander, D. C. (2003). Persistent angular structure: new insights fom diffusion magnetic resonance imaging data. *Inverse Problems*, 19:1031–1046.
- [Jbabdi et al., 2004] Jbabdi, S., Bellec, P., Marrelec, G., Perlberg, V., and Benali, H. (2004). A level set method for building anatomical connectivity paths between brain areas using DTI. In *International Symposium on Biomedical Imaging*, pages 1024–1027, Washinton DC, USA.
- [Jbabdi et al., 2007] Jbabdi, S., Woolrich, M., Andersson, J., and Behrens, T. (2007). A bayesian framework for global tractography. *NeuroImage*, 37:116–129.
- [Jeurissen et al., 2010] Jeurissen, B., Leemans, A., Jones, D. K., Tournier, J. D., and Sijbers, J. (2010). Probabilistic fiber tracking using the residual bootstrap with constrained spherical deconvolution. *Human Brain Mapping*.
- [Jeurissen et al., 2009] Jeurissen, B., Leemans, A., Tournier, J. D., and Sijbers, J. (2009). Fiber tracking on the 'fiber cup phantom' using constrained spherical deconvolution. In *MICCAI workshop on Diffusion Modelling and the Fiber Cup (DMFC'09)*, London, United Kingdom.
- [Jian and Vemuri, 2007] Jian, B. and Vemuri, B. C. (2007). Multi-fiber reconstruction from diffusion MRI using mixture of wisharts and sparse deconvolution. In *International Conference on Information Processing in Medical Imaging (IPMI)*.
- [Johansen-Berg and Rushworth, 2009] Johansen-Berg, H. and Rushworth, M. (2009). Using diffusion imaging to study human connectional anatomy. *The Lancet Neurology*, 32:75–94.
- [Kaden et al., 2007] Kaden, E., Knosche, T. R., and Anwander, A. (2007). Parametric spherical deconvolution: Inferring anatomical connectivity using diffusion MR imaging. *NeuroImage*, 37:474–488.
- [Kaufman et al., 1989] Kaufman, L., Kramer, D., Crooks, L., and Ortendahl, D. (1989). Measuring signal-to-noise ratios in mr imaging. *Radiology*, 173:265–267.
- [Klein et al., 2009] Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, L., Hellier, P., Song, J. H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T.,

- Woods, R. P., Mann, J. J., and Parsey, R. V. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*.
- [Koch et al., 2002] Koch, M., Norris, D., and Hund-Georgiadis, M. (2002). An investigation of functional and anatomical connectivity using magnetic resonance imaging. *NeuroImage*, 16:241–250.
- [Kreher et al., 2008] Kreher, B., Madeer, I., and Kiselev, V. (2008). Gibbs tracking: A novel approach for the reconstruction of neuronal pathways. *MRM*, 60:953–963.
- [Kreher et al., 2005] Kreher, B. W., Schneider, J. F., Mader, J., Martin, E., J. H., and Il’yasov, K. (2005). Multitensor approach for analysis and tracking of complex fiber configurations. *Magnetic Resonance in Medicine*, 54:1216–1225.
- [Lazar and Alexander, 2005] Lazar, M. and Alexander, A. L. (2005). Bootstrap white matter tractography (boot-tract). *NeuroImage*, 24:524–532.
- [Lazar et al., 2003] Lazar, M., Weinstein, D., Tsuruda, J., Hasan, K., Arfanakis, K., Meyerand, M., Badie, B., Rowley, H., V. Haughton, Field, A., and Alexander, A. (2003). White matter tractography using diffusion tensor deflection. In *Human Brain Mapping*, volume 18, pages 306–321.
- [Leemans et al., 2009] Leemans, A., Jeurissen, B., Sijbers, J., D.K., and Jones (2009). Exploredti: a graphical toolbox for processing, analyzing, and visualizing diffusion mr data. In *Proc. of the 17th Annual Meeting of Intl Soc Mag Reson Med*.
- [Leergaard et al., 2010] Leergaard, T. B., White, N. S., de Crespigny, A., Bolstad, I., D’Arceuil, H., Bjaalie, J. G., and Dale, A. M. (2010). Quantitative histological validation of diffusion mri fiber orientation distributions in the rat brain. *PloS one*, 5(1).
- [Lenglet, 2006] Lenglet, C. (2006). *Geometric and Variational Methods for Diffusion Tensor MRI Processing*. PhD thesis, Université de Nice-Sophia Antipolis.
- [Liu et al., 2004] Liu, C., Bammer, R., Acar, B., and Moseley, M. E. (2004). Characterizing non-gaussian diffusion by using generalized diffusion tensors. *Magnetic Resonance in Medicine*, 51:924–937.
- [Malcolm et al., 2009] Malcolm, J. G., Shenton, M. E., and Rathi, Y. (2009). Filtered tractography: Validation on a physical phantom. In *MICCAI workshop on Diffusion Modelling and the Fiber Cup (DMFC’09)*, London, United Kingdom.
- [Malcolm et al., 2010] Malcolm, J. G., Shenton, M. E., and Rathi, Y. (2010). Filtered multi-tensor tractography. *IEEE Trans. on Medical Imaging*, 29:1664–1675.
- [Mangin et al., 2002] Mangin, J.-F., Poupon, C., Cointepas, Y., Rivière, D., Papadopoulos-Orfanos, D., Clark, C. A., Régis, J., and Le Bihan, D. (2002). A framework based on spin glass models for the inference of anatomical connectivity from diffusion-weighted MR data. *NMR in Biomedicine*, 15:481–492.
- [Mori et al., 1999a] Mori, S., Crain, B., Chacko, V., and van Zijl, P. (1999a). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45:265–269.
- [Mori et al., 1999b] Mori, S., Crain, B., Chacko, V., and Zijl, P. V. (1999b). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Annals of Neurology*, 45:265–269.

- [Moussavi-Biugui et al., 2010] Moussavi-Biugui, A., Stieltjes, B., Fritzsche, K., Semmler, W., and Laun, F. B. (2010). Novel spherical phantoms for q-ball imaging under in vivo conditions. *Magnetic Resonance in Medicine*. In press.
- [Özarslan et al., 2006] Özarslan, E., Shepherd, T., Vemuri, B., Blackband, S., and Mareci, T. (2006). Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT). *NeuroImage*, 31(3):1086–1103.
- [Parker and Alexander, 2005] Parker, G. J. M. and Alexander, D. C. (2005). Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. *Philosophical Transactions of the Royal Society, Series B*, 360:893–902.
- [Parker et al., 2003] Parker, G. J. M., Haroon, H. A., and Wheeler-Kingshott, C. A. M. (2003). A framework for streamline-based probabilistic index of connectivity (PICO) using structural interpretation of MRI diffusion measurements. *Journal Magnetic Resonance Imaging*, 18:242–254.
- [Perrin et al., 2005a] Perrin, M., Poupon, C., Cointepas, Y., Rieul, B., Golestani, N., Pallier, C., Riviere, D., Constantinesco, A., Bihan, D. L., and Mangin, J.-F. (2005a). Fiber tracking in q-ball fields using regularized particle trajectories. In *Information Processing in Medical Imaging*, pages 52–63.
- [Perrin et al., 2005b] Perrin, M., Poupon, C., Cointepas, Y., Rieul, B., Golestani, N., Pallier, C., Rivière, D., Constantinesco, A., Bihan, D. L., and Mangin, J.-F. (2005b). Fiber tracking in q-ball fields using regularized particle trajectories. In *IPMI*, pages 52–63.
- [Pieper et al., 2006] Pieper, S., Lorensen, B., Schroeder, W., and Kikinis, R. (2006). The NA-MIC kit: ITK, VTK, pipelines, grids and 3D Slicer as an open platform for the medical image computing community. In *Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, volume 1, pages 698–701.
- [Poupon et al., 2010] Poupon, C., Guevara, P., Larivière, L., Tournier, G., Bernard, J., Fillard, P., Descoteaux, M., and Mangin, J.-F. (2010). A human-like coronal tractography phantom. In *In proceedings of the joint ISMRM/ESMRMB conference*.
- [Poupon et al., 2001] Poupon, C., Mangin, J.-F., Clark, C., Frouin, V., Régis, J., LeBihan, D., and Bloch, I. (2001). Toward inference of human brain connectivity from MR diffusion tensor data. *MedIA*, 5(1):1–15.
- [Poupon et al., 2008] Poupon, C., Rieul, B., Kezele, I., Perrin, M., Poupon, F., and Mangin, J.-F. (2008). New diffusion phantoms dedicated to the study and validation of HARDI models. *Magnetic Resonance in Medicine*, 60:1276–1283.
- [Pullens et al., 2010] Pullens, P., Roebroek, A., and Goebel, R. (2010). Ground truth hardware phantoms for validation of diffusion-weighted mri applications. *Journal of Magnetic Resonance Imaging*, 32:482–488.
- [Ramirez-Manzanares and Rivera, 2006] Ramirez-Manzanares, A. and Rivera, M. (2006). Basis tensor decomposition for restoring intra-voxel structure and stochastic walks for inferring brain connectivity in DT-MRI. *International Journal of Computer Vision*, 69(1):77 – 92.

- [Ramirez-Manzanares et al., 2009] Ramirez-Manzanares, A., Rivera, M., and Gee, J. C. (2009). Depicting axon fibers on a diffusion phantom by means of hybrid dbf-dt data. In *MICCAI workshop on Diffusion Modelling and the Fiber Cup (DMFC'09)*, London, United Kingdom.
- [Ramirez-Manzanares et al., 2007] Ramirez-Manzanares, A., Rivera, M., Vemuri, B., Carney, P., and Mareci, T. (2007). Diffusion basis functions decomposition for estimating white matter intravoxel fiber geometry. *IEEE Transactions on Medical Imaging*, 26(8):1091–1102.
- [Reisert et al., 2009] Reisert, M., Mader, I., and Kiselev, V. (2009). Tracking a physical phantom by global fibre reconstruction. In *MICCAI workshop on Diffusion Modelling and the Fiber Cup (DMFC'09)*, London, United Kingdom.
- [Rivière et al., 2009] Rivière, D., Geffroy, D., Denghien, I., Souedet, N., and Coin-tepas, Y. (2009). BrainVISA: an extensible software environment for sharing multimodal neuroimaging data and processing tools. In *Proc. 15th HBM*.
- [Sakaie, 2009] Sakaie, K. (2009). Fast persistent angular structure based streamline tractography. In *MICCAI workshop on Diffusion Modelling and the Fiber Cup (DMFC'09)*, London, United Kingdom.
- [Savadjiev et al., 2008] Savadjiev, P., Campbell, J. S. W., Descoteaux, M., Deriche, R., Pike, G. B., and Siddiqi, K. (2008). Labeling of ambiguous sub-voxel fibre bundle configurations in high angular resolution diffusion MRI. *NeuroImage*, 41(1):58–68.
- [Seunarine et al., 2006] Seunarine, K. K., Cook, P. A., Embleton, K., Parker, G. J. M., and Alexander, D. C. (2006). A general framework for multiple-fibre PICO tractography. In *Medical Image Understanding and Analysis*.
- [Seunarine et al., 2007] Seunarine, K. K., Cook, P. A., Hall, M. G., Embleton, K. V., Parker, G. J. M., and Alexander, D. C. (2007). Exploiting peak anisotropy for tracking through complex structures. In *Mathematical Methods in Biomedical Image Analysis (MMBIA 2007), held in conjuncture with the 11th International Conference on Computer Vision*, pages 1–8.
- [Tensaouti et al., 2009] Tensaouti, F., Lotterie, J. A., and Berry, I. (2009). Fiber tracking on the phantom dataset by using sisyph software. In *MICCAI workshop on Diffusion Modelling and the Fiber Cup (DMFC'09)*, London, United Kingdom.
- [Tournier et al., 2007] Tournier, J.-D., Calamante, F., and Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion mri: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, 35(4):1459–1472.
- [Tournier et al., 2004] Tournier, J.-D., Calamante, F., Gadian, D., and Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, 23:1176–1185.
- [Toussaint et al., 2007] Toussaint, N., Souplet, J., and Fillard, P. (2007). Medinria: Medical image navigation and research tool by inria. In *Proc. of MICCAI'07 Workshop on Interaction in medical image analysis and visualization*, Brisbane, Australia.
- [Tuch, 2004] Tuch, D. (2004). Q-ball imaging. *Magnetic Resonance in Medicine*, 52(6):1358–1372.

- [Tuch, 2002] Tuch, D. S. (2002). *Diffusion MRI of Complex Tissue Structure*. PhD thesis, Harvard University and Massachusetts Institute of Technology.
- [Wang et al., 2007] Wang, R., Benner, T., Sorensen, A. G., and Wedeen, V. J. (2007). Diffusion toolkit: A software package for diffusion imaging data processing and tractography. In *Proceedings of the 15th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*.
- [Wang and Wedeen, 2007] Wang, R. and Wedeen, V. (2007). A software package for diffusion imaging data processing and tractography. In *Proceedings of the 15th Scientific Meeting of the International Society for Magnetic Resonance in Medicine*.
- [Weinstein et al., 1999] Weinstein, D. M., Kindlmann, G. L., and Lundberg, E. C. (1999). Tensorlines: Advection-diffusion based propagation through diffusion tensor fields. *IEEE Vis.*, 00:40.
- [Yo et al., 2009] Yo, T.-S., Anwender, A., and Knäusche, T. R. (2009). Fiber cup 2009: Reconstructing fibers from the phantom data. In *MICCAI workshop on Diffusion Modelling and the Fiber Cup (DMFC'09)*, London, United Kingdom.
- [Zhang et al., 2009] Zhang, F., Hancock, E. R., Goodlett, C., and Gerig, G. (2009). Probabilistic white matter fiber tracking using particle filtering and von mises-fisher sampling. *Medical Image Analysis*, 13(1):5–18.